

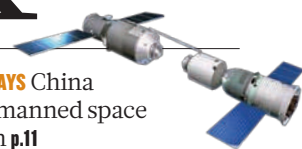
THIS WEEK

EDITORIALS

GALILEO Why the great man should see Jupiter up close **p.6**

WORLD VIEW How a climate scientist took on a newspaper and won **p.7**

SEVEN DAYS China plans manned space station **p.11**



Support refugee scientists

The cause of displaced scholars provides a much-needed reminder that intellectual freedom must not be taken for granted. Groups that help them need greater support themselves.

Albie Sachs, a white South African human-rights lawyer, was imprisoned by the apartheid regime in the 1960s and lost an arm and was blinded in one eye in 1988, in a bomb assassination attempt by security forces. On each occasion, he found a safe haven as an academic in Britain from which to rebound. After the end of apartheid, Sachs returned to South Africa, where he is credited as an architect of the country's progressive post-apartheid constitution, and sat on its Constitutional Court from 1994 until last year.

Sachs, like thousands of other refugees over the years, received a lifeline from the Council for Assisting Refugee Academics (CARA), a small charity based at the London South Bank University that helps academics and scientists who are seeking, or have obtained, political asylum in the United Kingdom to make a fresh start. Working with a growing network of universities in Britain — and increasingly in continental Europe — it assists refugees in finding academic placements. It also provides small grants, often just a couple of thousand pounds, to help them settle in, finish degrees or requalify. CARA offered “a little grant at the right moment, but it was much more than a grant”, Sachs said. What counted most was “the human connection”.

VIGILANCE NEEDED

In much of the free world, a threat to academic freedom usually means political interference in science-based policies, or pressure for research to show an economic pay-off. But in many parts of the world, academic freedom has an altogether starker meaning — academics and their families can face discrimination, prison, or worse, for speaking out about or studying issues that threaten dominant policies or ideologies. They can also be persecuted for their politics, or for belonging to a particular ethnic group.

Helping refugee academics is worthy in itself, and raises scientists' awareness of the need for vigilance in protecting academic freedom. That message is all too relevant today. In an increasingly individualistic and competitive research world, academics tend to be preoccupied with their own interests. At the same time, tough anti-immigration stances are taking hold in several European countries. Political asylum is increasingly difficult to obtain in countries including France and the United Kingdom, and refugee academics from Islamic countries often face hostility and discrimination.

History shows that no country is immune to political and ideological changes that can repress intellectual freedom, warned Rowan Williams, the Archbishop of Canterbury, in the CARA annual lecture earlier this year. He succinctly summed up what is at stake: “Defending intellectual freedom is defending the possibility not only of a free academy but of a society willing to learn — and thus a society willing to see itself critically.”

Founded in the 1930s to help scientists in continental Europe fleeing the Nazis and other fascist regimes, CARA supported

some 1,500 academics in those dark years, 16 of whom went on to win Nobel prizes. It currently aids around 200 refugee academics annually. Other agencies worldwide have also taken up the cause, including the US-based Scholars at Risk and the Scholar Rescue Fund. Some learned societies, such as the American Physical Society and the American Association for the Advancement of Science, are also actively publicizing cases of academics under threat. But the integration of refugee scholars needs a firmer footing and better funding — CARA has eight paid staff and an annual budget of some £840,000 (US\$1.2 million) — as well as more institutions to play an active host role.

“Supporting beleaguered academics is key to defending the aspirations of open civil societies.”

The effort also faces new challenges. Many of those seeking CARA's aid are academics from poorer countries. Often, although they held senior positions at home, they may not be qualified to compete in the cut-throat academic environment of richer countries.

In response, CARA often tries to deploy such scientists on research relevant to their home regions, for which they can be well qualified. Or it helps them to find jobs outside academia.

In some cases, aiding scientists in their home country might make more sense. CARA maintains a strong focus on refugee academics from Iraq, for example, arguing that violence continues and noting that two researchers who recently returned to Iraq were killed. But although academics remain at risk in Iraq, they no longer seem to be targeted on a large scale (see *Nature* **441**, 1036–1037; 2006). It could be argued that the greater need now is for them to stay in Iraq when possible, to help in the huge task of reconstructing the country's gutted research and higher-education system.

REACHING OUT

CARA has taken steps in this direction, for example by broadening its initial mission to include funding fellowships outside the United Kingdom. In the Middle East, CARA supports research on topics such as tuberculosis, education and depleted-uranium reduction, with researchers at UK universities as unpaid collaborators. In Zimbabwe, CARA is also in the early stages of a project to help rebuild the tattered higher-education system, by funding lecture trips for expatriate Zimbabwean academics, encouraging others to return, and setting up high-speed Internet connections at the country's universities.

Governments and other funders would do well to consider contributing to an expansion of such efforts. Supporting beleaguered academics preserves talent, and is key to defending the aspirations of open civil societies. Due diligence is needed to ensure that the money is well spent. But given the human return on investment, such activities are a bargain. ■

Galileo's send-off

An upcoming mission to Jupiter should include a piece of the famous astronomer.

At an aerospace facility in Denver, Colorado, engineers are busy attaching scientific instruments to NASA's next mission to Jupiter, which is set for launch in less than a year. But team members on the billion-dollar Juno mission are quietly talking about slipping something extra onto the spacecraft — a tiny fragment of bone from Galileo Galilei.

The idea of sending a piece of the famed astronomer to orbit the giant planet, in the company of the moons that he discovered, has charmed some of the US participants in the mission. Officials at the Italian Space Agency, which is providing two instruments, seem to be less enthusiastic. But the plan should move forward.

The wariness on the Italian side is understandable. Galileo died in 1642, still under house arrest for his heretical views, and his burial place lay hidden until 1737, when civil authorities moved him to a prominent spot in the Basilica of Santa Croce in Florence, Italy. Since then, the Franciscan order that runs the church has done its best to protect his remains. A 2008 proposal to exhume the body for genetic tests met with fierce opposition from the church's rector, who described such high-profile scientific efforts as resembling "a carnival".

Yet Galileo has never been wholly at rest. During the reburial,

attending naturalist Giovanni Targioni Tozzetti pried off the thumb and forefinger of Galileo's right hand for posterity; earlier this year, the bones went on display at a museum a few blocks from Santa Croce. And one of the astronomer's vertebrae is held by the University of Padua, where he once taught.

"A sliver of bone seems a modest sacrifice for a gesture that would remind the public that science is a human endeavour."

Compared with these disruptions of Galileo's eternal slumber, a sliver of bone seems a modest sacrifice for a gesture that would add emotional energy to the mission and remind the public that science is fundamentally a human endeavour.

Galileo himself understood the need to connect science with society, and he was politically astute enough to see the value of showmanship. It was no coincidence that he named the four biggest moons of Jupiter the 'Medicean stars' after his patrons, and accompanied his announcement of their discovery with a gift of an almost priceless telescope to Cosimo II de' Medici, the Grand Duke of Tuscany. Galileo might have tolerated a little post-mortem exploitation in the interests of public support.

But there is another reason to think that Galileo would have approved. He was enthralled by the Universe revealed through his telescope. "It is a very beautiful thing, and most gratifying to the sight, to behold the body of the moon, distant from us almost sixty earthly radii, as if it were no farther away than two such measures," he wrote of the lunar surface in his 1610 treatise *The Starry Messenger*. The Juno mission will skim just 4,800 kilometres above Jupiter, a body that held no less fascination for Galileo. He just might enjoy a closer look. ■

A painful remedy

The number of papers being retracted is on the rise, for reasons that are not all bad.

Few experiences can be more painful to a researcher than having to retract a research paper. Some papers die quietly, such as when other scientists find that the work cannot be replicated and simply ignore it. Yet, as highlighted by several episodes in recent years, the most excruciating revelation must be to find not only that a paper is wrong, but that it is the result of fraud or fabrication, which itself requires months or years of investigation. Where once the research seemed something to be exceptionally proud of, the damage caused by fraudulent work can spread much wider, as discovered by associates of the Austrian physicist Jan Hendrick Schön and the South Korean stem-cell biologist Woo Suk Hwang. But whatever the reason for a retraction, all of the parties involved — journals included — need to face up to it promptly.

This year, *Nature* has published four retractions, an unusually large number. In 2009 we published one. Throughout the past decade, we have averaged about two per year, compared with about one per year in the 1990s, excluding the pulse of retractions of papers co-authored by Schön.

Given that *Nature* publishes about 800 papers a year, the total is not particularly alarming, especially because only some of the retractions are due to proven misconduct. A few of the *Nature* research journals have also had to retract papers in recent years, but the combined data do no more than hint at a trend. A broader survey revealed even smaller proportions: in 2009, *Times Higher Education* commissioned a survey by Thomson Reuters that counted 95 retractions among 1.4 million papers published in 2008. But the same survey showed that, since 1990 — during which time the number of published papers doubled — the proportion of retractions increased tenfold (see go.nature.com/vphd17).

So why the increase? More awareness of misconduct by journals

and the community, an increased ability to create and to detect unduly manipulated images, and greater willingness by journals to publish retractions must account for some of this rise. One can also speculate about the increasing difficulty for senior researchers of keeping track of the detail of what is happening in their labs. This is of concern not just because of the rare instances of misconduct, but also because of the risk of sloppiness and of errors not being caught. Any lab with more than ten researchers may need to take special measures if a principal investigator is to be able to assure the quality of junior members' work.

The need for quality assurance and the difficulties of doing it are exacerbated when new techniques are rapidly taken up within what is often a highly competitive community. And past episodes have shown the risk that collaborating scientists — especially those who are geographically distant — may fail to check data from other labs for which, as co-authors, they are ultimately responsible.

If we at *Nature* are alerted to possibly false results by somebody who was not an author of the original paper, we will investigate. This is true even if the allegations are anonymous — some important retractions in the literature have arisen from anonymous whistle-blowing. However, we are well aware of the great damage that can be done to co-authors as a result of such allegations, especially when the claims turn out to be false. Such was the case with a recent e-mail alert widely distributed by a group calling itself Stem Cell Watch (see *Nature* 467, 1020; 2010) — an action that we deplore.

For our part, we are sensitive to such concerns and will bear in mind the need to protect the interests of authors until our obligation to the community at large becomes clear. But then we will publish a retraction promptly, and link to it prominently from the original papers. We will also list the retraction on our press release if the original paper was itself highlighted to the media.

Ultimately, it comes down to the researchers — those most affected by the acts — to remain observant and diligent in pursuing their concerns wherever they lead, and where necessary, to correct the literature promptly. Too often, such conscientious behaviour is not rewarded as it should be. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunq



How to beat the media in the climate street fight

Researchers must take a more aggressive approach to counter shoddy journalism and set the scientific record straight, says **Simon L. Lewis**.

When science hits the news, researchers often moan about the quality of the coverage. A sharp reminder of the issue rolls round this month — the anniversary of the global media frenzy over the release of e-mails from climate researchers at the University of East Anglia, UK. So what should scientists do when reporting quality falls off a cliff? Earlier this year, I was seriously misrepresented by a newspaper and thrown into a political storm. Rather than take it lying down, I set the record straight. It has been an odd journey, and I think there are lessons for how we scientists should deal with the media.

In January, the absurd claim from the Intergovernmental Panel on Climate Change (IPCC) that Himalayan glaciers would melt by 2035 launched a hunt to find other exaggerated risks of climate change. A British blogger, Richard North, found an IPCC statement that part of the Amazon rainforest may be at risk from droughts, referenced to an environment group's report, not the scientific literature. North dubbed it Amazongate, and told the world that the IPCC view "seems to be a complete fabrication".

As a tropical-forest expert, I found my telephone ringing for three days. Journalists asked me to comment on the IPCC line that "up to 40% of the Amazonian forests could react drastically to even a slight reduction in precipitation". My short answer was that in context, the statement was broadly correct; but the wording was not careful, and the IPCC should have cited the primary literature. My comments were broadcast across the BBC, but for most news outlets it was a non-story.

The Sunday Times saw it differently. Its reporter, Jonathan Leake, asked both leading and genuinely inquisitive questions. I sent him scientific papers, and we discussed them. He agreed to read the finished piece to me over the telephone before publication. It stated, correctly, that the future of the Amazon is very uncertain, because the available data are limited. I was quietly pleased that I had 'spun' what I saw as a blogger's anti-IPCC tirade into a story about the science. Yet I was wrong. The newspaper headline was "UN climate panel shamed by bogus rainforest claim", and worse, I was the expert quoted to support it. The article had been completely rewritten, essentially parroting North's blog, to include new quotes from me (genuine, but heavily edited and misleadingly taken out of context), and fabricated assertions about my views. An accompanying editorial called for the IPCC chairman to resign.

I was furious. Worse, the two conflicting versions of my views — on the BBC and in *The Sunday Times* — constituted a serious affront to my professional credibility. But what could I do? I added a comment under the online version of the article that my views were not accurately reported, and sent a letter for publication to *The Sunday Times*.

Weeks later the misleading article had been reproduced over 20,000 times on the Internet. My letter had been ignored and website comment deleted. Furthermore, my words and standing as an expert were being used by other newspapers to allege widespread corruption by IPCC scientists. As an Editorial on climate disinformation in this journal said at the time: "Scientists must now emphasize the science, while acknowledging that they are in a street fight." I needed to fight back.

After advice from a friend in public relations and press officers at scientific organizations, I filed an official complaint to the Press Complaints Commission, the UK media watchdog. The commission could order the newspaper to print a correction, but would that happen and was it enough? I needed to make the complaint itself a story.

I contacted *The Guardian* newspaper, which published an article about my complaint. To reach the US audience, I handed the full complaint as an exclusive to perhaps the world's most influential political climate-change blog, Joe Romm's climateprogress.org.

For a scientist to take such an active media role was unorthodox, but it felt good. And it worked. It was widely recognized that the story was wrong and I had been badly treated. *The New York Times* featured me in a front-page article.

The Sunday Times offered to publish a single-line apology. I knew others had extracted greater concessions and kicked harder. It eventually agreed to remove the article from its website, and replace it with a formal correction and apology, also printed prominently in the newspaper. The retraction was reported around the world.

Environmental commentators hailed the apology as vindication for the IPCC (which it wasn't quite, as its statements were not faultless). Climate sceptics launched a counter-attack by claiming that no apology was due because the IPCC statement was not perfect. But for me the storm had passed.

What lessons are there for scientists in politically charged areas who find themselves in a similar position? Do your research. What is the reporter's track record? Anticipate that every sentence you say or write may be dissected and interpreted in the least charitable manner possible. And if things go wrong, seek advice from public-relations experts, and where necessary, media lawyers. In my experience, science-media professionals are almost as lost as scientists themselves, when dealing with topics as emotive as climate change.

The media dictate what most people know about contemporary scientific debates. Given the need for informed policy, scientists need to learn to better read and engage with this media landscape. Closing the newspaper with a sigh is not enough. ■

Simon L. Lewis is a Royal Society research fellow and reader in global change science at the University of Leeds.
e-mail: s.l.lewis@leeds.ac.uk

**CLOSING
THE NEWSPAPER
WITH A
SIGH
IS NOT ENOUGH.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/luu1tc

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

CHEMISTRY

Zinc can swim in an electric field

A tiny zinc object that can propel itself from one end of a capillary tube to the other holds promise for those striving to make minuscule motors.

Gabriel Loget and Alexander Kuhn at the University of Bordeaux in France immersed their micrometre-sized zinc 'swimmer' in a solution of zinc sulphate and applied a relatively weak external electric field. The field promotes oxidation on one side of the swimmer, dissolving the zinc, and reduction on the other, which deposits zinc, thus moving the object along. The swimmer reached speeds of 80 micrometres per second.

The authors suggest that their method could be an efficient way to control the motion of metallic micro-objects in engineered micromotors — or perhaps even nanomotors.

J. Am. Chem. Soc. doi:10.1021/ja107644x (2010)

GLACIOLOGY

Heat trickles down in ice sheets

Meltwater is thought to be contributing to accelerated glacier flow. Thomas Phillips and his colleagues at the University of Colorado at

Boulder show that meltwater seeping through fractures and crevasses in the ice (pictured) can have a sustained warming effect, speeding up ice flow.

By incorporating the effect of this seeping water into thermal models of ice sheets, the authors found that the relatively temperate

meltwater warms the surrounding ice, thereby reducing its viscosity and helping to increase its flow. Most of the water refreezes during the winter, but small quantities can remain liquid, causing the sustained warming.

This effect could be leading to

much faster glacial responses to warming than current models predict. It is probably contributing to observed changes in some regions of the Greenland ice sheet, the authors suggest.

Geophys. Res. Lett. doi:10.1029/2010GL044397 (2010)

CANCER BIOLOGY

Cell neighbours aid cancer relapse

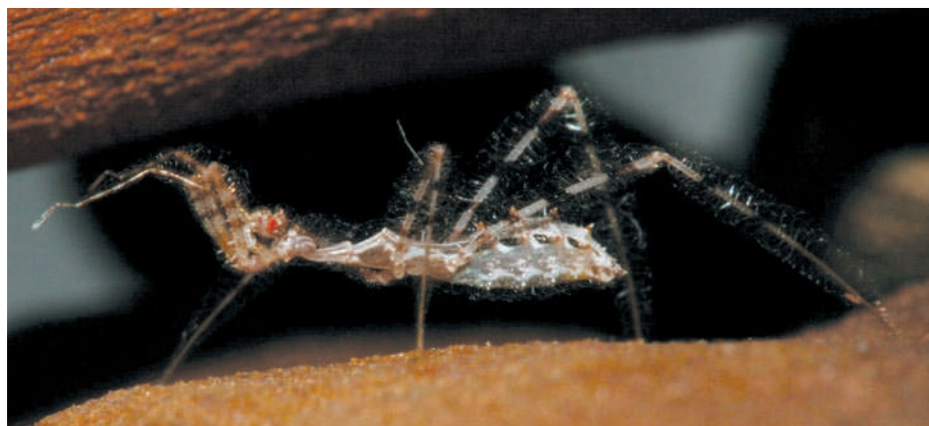
Cancer patients often relapse after seemingly successful chemotherapy — perhaps because cellular stress responses to the drugs shield some cancer cells, creating a protected reservoir that can seed fresh tumour growth.

Luke Gilbert and Michael

Hemann at the Massachusetts Institute of Technology in Cambridge studied the environment surrounding tumour cells in a mouse model of Burkitt's lymphoma, a cancer of the lymphatic system. They found that normal cells in the thymus, an organ that consists mainly of lymphatic tissue, release two proteins in response to the DNA damage caused by chemotherapy. Those proteins, IL-6 and Timp-1, then protect nearby lymphoma cells from the lethal effects of the drugs.

The same thing happened in human liver-cancer cells grown in culture and treated with doxorubicin, a common treatment for liver cancer.

Cell 143, 355–366 (2010)



A. WIGNALL

ZOOLOGY

Spiders lured by subterfuge

Assassin bugs hunt spiders by plucking the threads of spiderwebs to draw the arachnids closer. Researchers in Australia report that the vibrations mimic those generated by insects trapped in the webs — and are even at specific frequencies to avoid eliciting an aggressive response.

In a series of laboratory tests, Anne Wignall and Phillip Taylor of Macquarie University in Sydney show that the vibrations induced by

Stenolemus bituberus (pictured) trigger the same behaviour from spiders as do those of some prey species. Vibrations from falling leaves elicited no response, whereas those from courting male spiders led females to adopt copulatory positions.

The assassin bugs seem to avoid producing the higher frequency vibrations generated by some prey that can trigger a dangerous high-speed approach from the spider.

Proc. R. Soc. B doi:10.1098/rspb.2010.2060 (2010)



K. STEFFEN

ECOLOGY

Brave fish forage more

Fishing can affect not just the numbers of predators and the behaviour of their prey, but also the distribution of plants at the bottom of the food chain.

Elizabeth Madin, currently at the University of Technology–Sydney in Australia, and her colleagues developed a model assuming that a reduction in predators from fishing makes prey species bolder. They are then inclined to forage farther afield.

The authors used the model to accurately predict the spatial distribution of seaweed, which is eaten by the prey fish, on the coral reefs of the northern Line Islands in the Pacific Ocean. Those of Palmyra Atoll, which has seen almost no fishing, had more and larger seaweed patches — indicating less adventurous foraging — than did the intensely fished reefs of Kiritimati Atoll.

Am. Nat. doi:10.1086/657039 (2010)

APPLIED PHYSICS

Long-lasting memory storage

A data-storage device based on silver ions embedded in glass could theoretically store 180 gigabytes per square centimetre — about 80 times the capacity of a Blu-ray disc.

Lionel Canioni at the University of Bordeaux in France and his team embedded a three-dimensional matrix of silver ions in light-sensitive glass. The ions fluoresce when hit with a laser. The researchers could control the silver's fluorescence at 16 discrete levels, coding information as varying levels of brightness rather than as the binary 'ones' and 'zeros' that conventional magnetic storage devices use.

The authors managed to store and retrieve micrometre-sized pictures of three French Nobel prizewinners. Because glass resists wear from temperature and humidity, the

authors say that it could be used for high-density information storage that lasts centuries.

Adv. Mater. doi:10.1002/adma.201002413 (2010)

IMMUNOLOGY

How NETs form to capture invaders

As a first line of defence against microbes, immune cells called neutrophils release webs of uncoiled DNA strands and antimicrobial molecules to ensnare and kill the invaders (**pictured**). Researchers have identified two enzymes that work together to help to form these traps, known as NETs.

Neutrophils generate reactive oxygen-containing molecules and then undergo cell death, releasing the NETs. Arturo Zychlinsky and his team at the Max Planck Institute for Infection Biology in Berlin tested neutrophil extracts for their ability to unwind DNA and disrupt the cell nucleus — key steps in NET formation.



They homed in on two enzymes, neutrophil elastase and myeloperoxidase, that are normally stored in granules. When activated by the oxygen-containing molecules, the enzymes move to the nucleus, where they unwind DNA. *J. Cell Biol.* doi:10.1083/jcb.201006052 (2010)

ECOLOGY

Climate controls tree growth

Climatic factors such as rainfall, seasonality and temperature have a greater

COMMUNITY CHOICE

The most viewed papers in science

GENETICS

The enigma of genes and environment

HIGHLY READ
on plogenetics.org
11–24 October

The interplay between an animal's genes and environment at the molecular level is largely mysterious. Researchers report that, even in a simple yeast model, the

two interact in ways that are difficult to predict.

Barak Cohen and his colleagues at Washington University in St Louis, Missouri, created 32 yeast strains carrying different combinations of four single-nucleotide gene variants that affect how efficiently the yeast forms spores. The strains were generated with two different genetic backgrounds and cultured in eight different growth media. The researchers then tracked changes in spore formation under each condition.

The authors were able to predict the effects of a genetic variant on spore formation, but only after they had accounted for a strain's growth environment and genetic background. This suggests that those studying human genetics and disease need to measure the effects of such factors.

PLoS Genet. 6, e1001144 (2010)

effect on tropical tree growth than logging or soil nutrient levels, a study of Bolivian lowland forests reveals.

Marisol Toledo of the Bolivian Forest Research Institute in Santa Cruz and her colleagues monitored tree growth in 165 one-hectare plots for periods of up to 11 years, and collected data on climate, soil nutrition and disturbance. They found that logging does increase growth, but to a lesser extent than climate, and the effect disappears six years after trees are felled. Nutrient levels in soil do not have a large effect.

Forest growth is projected to slow with climate change, as rainfall is predicted to decrease and seasonality to increase in the tropics. However, these changes could be partially offset by rising temperatures. *J. Ecol.* doi:10.1111/j.1365-2745.2010.01741.x (2010)

STRUCTURAL BIOLOGY

A walk through the genome

Genes are regulated in numerous ways; one is the addition of methyl groups to histone proteins, which

bind and package up DNA. A study of a viral enzyme that methylates histones suggests that it functions by 'walking' from one methylation site to the next.

The enzyme, called vSET, is found in *Paramecium bursaria* chlorella virus 1 and suppresses gene expression in the host to aid infection. Hua Wei and Ming-Ming Zhou at Mount Sinai School of Medicine in New York determined the three-dimensional structure of the enzyme and studied how it binds to histones. They surmised that vSET has two binding areas, of which only one can bind to the substrate at a time. Once methylation has occurred at the first binding site, vSET is released at the same time that its second binding area attaches to another site. This 'walking' mechanism allows the enzyme to work efficiently across the whole genome, the authors say. *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.1009911107 (2010)

► **NATURE.COM**

For the latest research published by Nature visit:

www.nature.com/latestresearch

SEVEN DAYS

The news in brief

BUSINESS

Industry R&D drop

Global investment in commercial research and development (R&D) by the world's biggest research spenders dropped by just 1.9% in 2009, according to the European Union's annual investment scoreboard, released on 26 October. That ends four consecutive years of growth in R&D, but the 1,400 companies in the survey experienced, on average, a 10.1% drop in sales and a 21% fall in profits in 2009. See go.nature.com/mv63xe for more.

Gene patent twist

The US government has weighed in on the court battle over patents on the genes *BRCA1* and *BRCA2*, which are the target of a proprietary test to assess the risk of breast cancer. A New York district court ruled in March that some aspects of the cancer-gene patents are invalid, and the case is now under appeal. The government argues that although modified DNA sequences should be patentable, unmodified DNA should not. This contradicts current practice at the US Patent and Trademark Office. See go.nature.com/q7xi1t for more.

Drug settlement

London-based pharmaceutical giant GlaxoSmithKline has agreed to plead guilty and pay \$750 million in criminal and civil fines for producing adulterated drugs, the US Department of Justice announced on 26 October. The fines cover offences committed at a plant in Puerto Rico that closed in 2009. They include failure to ensure that an anti-nausea drug was free of microbial contamination; the production of a diabetes drug that contained too much

or too little active ingredient; and the manufacture of two-layer tablets of the antidepressant paroxetine that split, leaving some without therapeutic effect.

Sequence squabble

Pacific Biosciences, a DNA-sequencing company based in Menlo Park, California, completed its initial public offering on 26 October, raising US\$200 million after emerging onto the market with 12.5 million shares sold for \$16 each. Meanwhile, Helicos BioSciences of Cambridge, Massachusetts, expanded its patent infringement lawsuit against Pacific Biosciences to include Life Technologies in Carlsbad and Illumina in San Diego, two other California-based sequencing firms. Helicos, which is threatened

with delisting from the NASDAQ market, claims that it was the first to invent single-molecule sequencing, on which many new technologies in development now rely.

POLICY

Biodiversity deal

Last week's negotiations on the Convention on Biological Diversity in Nagoya, Japan, delivered a new set of conservation targets for 2020, and an additional treaty that lays down rules for sharing the commercial benefits of products derived from the world's flora and fauna. See page 14 for more.

Nuclear storage

A European Commission directive expected to be published this week proposes

that nuclear waste should be stored deep underground. The draft law says that it is "the safest and most sustainable" way of storing spent fuel and other waste over the long term. France, Finland and Sweden are already planning to open underground nuclear-waste repositories within the next 20 years (see *Nature* **466**, 804–805; 2010).

NIH rules rankle

Scientists at the US National Institutes of Health (NIH) work less with industry as a result of a 2005 crackdown that barred employees from consulting for pharmaceutical companies and other businesses. A survey of 566 agency scientists, published in the November issue of *Academic Medicine*, finds that just 33% maintain ties with industry, compared



REUTERS/E. QUEIROZ/A. CRITICA

Brazil's drought worsens

The Amazon basin is facing its worst drought since 2005. Water levels in the Rio Negro, a primary tributary, are at historic lows and parts of the river are unnavigable. The river was at its highest levels in June 2009, owing to heavy rains.

It is hard to pinpoint a culprit, but both the 2005 and the 2010 droughts align well with longer-term projections by some climate modellers for a drying out of the Amazon due to global warming. See go.nature.com/mhffyl for more.

with 52% before the rules were implemented, even though they are publishing just as many papers. Three-quarters say that the rules have impeded the NIH's scientific mission. See go.nature.com/qwssjn for more.

Stem-cell funding

The US Court of Appeals for the District of Columbia Circuit has set 6 December as the date it will hear oral arguments in a lawsuit challenging the legality of the US government funding human embryonic stem-cell research. The government is appealing against an injunction issued in August by a lower court, which shut down National Institutes of Health funding for the controversial research on the grounds that it contravenes an existing law that bans federal funding for research in which embryos are destroyed. See www.nature.com/stemcellfunding for more.

Not so NICE

Uncertainty emerged last week over the future of Britain's National Institute for Health and Clinical Excellence (NICE), which assesses whether drugs are sufficiently cost-effective to be made available through the country's National Health Service. The government plans to introduce a value-based pricing scheme

in 2013, negotiating fees with manufacturers on the basis of a drug's clinical value. NICE will have an advisory role in the process. Analysts disagree on whether the pricing mechanism will help or hinder pharmaceutical research and development. See go.nature.com/vsguxm for more.

Brazilian election

Dilma Rousseff was elected as Brazil's new president on 31 October. Nominated by incumbent President Luiz Inácio Lula da Silva as his successor, she is widely expected to continue his policy of increasing investment in science (see *Nature* **467**, 511; 2010).

FUNDING

Rouble roll-out

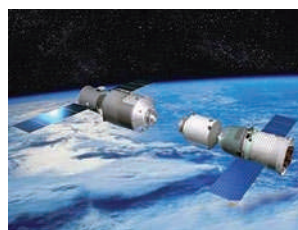
The Russian Federation has chosen 40 scientists to receive 'megagrants' worth up to US\$4.9 million each, as part of efforts to build up government-supported science (see *Nature* **467**, 251; 2010). More than 500 researchers applied, including many non-Russians who agreed to spend four months a year working in a Russian lab to qualify. Winners include Ferid Murad, a biochemist at the University of Texas in Houston and co-recipient of the 1998 Nobel Prize in Physiology or Medicine, and Stanislav Smirnov, a mathematician

at the University of Geneva, Switzerland, who won the Fields Medal this year. A second funding round will select another 40 recipients in spring 2011.

RESEARCH

Space station

China announced official plans for a manned space station on 27 October. The first section, an 8.5-tonne unmanned module, will go into orbit in early 2011, and will dock with another unmanned space vehicle scheduled for launch later that year (**pictured**). The first astronauts will arrive a year later, bringing additional components, and a space laboratory is scheduled to be in place by 2016. The station should be finished by 2020, two years ahead of earlier projections.



Fraud claim wrong

An investigation by *Nature* has found no sign of fraud in a stem-cell paper from Konrad Hochedlinger's group at the Harvard Stem Cell Institute (J. Utikal *et al.* *Nature*

COMING UP

8-9 NOVEMBER

The Royal Society holds a meeting in London on the status of geoengineering. go.nature.com/hhtf3f

8-12 NOVEMBER

Delegates to the Montreal Protocol for curbing ozone-depleting substances meet in Bangkok. They may decide to regulate hydrofluorocarbons — refrigerants that are ozone-friendly but are potent greenhouse gases. go.nature.com/a9s1wl

9 NOVEMBER

The International Energy Agency releases *World Energy Outlook 2010* in London, with recommendations on how to meet the 2009 Copenhagen Accord, which outlined aspirational goals for tackling climate change. go.nature.com/orprbp

460, 1145–1148; 2009). In October, an anonymous group calling itself Stem Cell Watch accused the paper's authors of inappropriately manipulating an image in the paper (see *Nature* **467**, 1020; 2010).

Biobank planned

Scientists from Qatar and Britain last week announced plans to launch what they say is the first large population-based biomedical study in an Arabic-speaking nation. The Qatar Biobank will collect DNA, blood and other biological samples from up to 100,000 volunteers. Scientists will use the biobank to uncover markers linked to conditions such as diabetes and heart disease, which are on the rise among Qataris.

➔ NATURE.COM

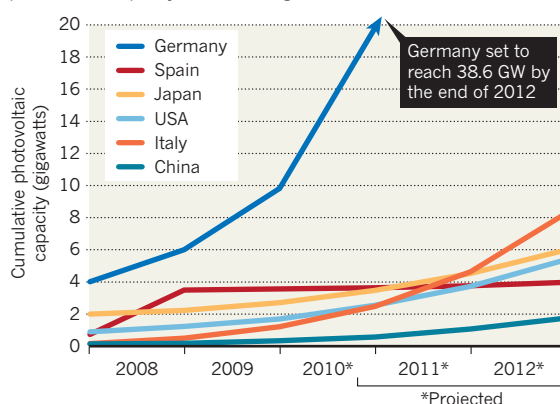
For daily news updates see: www.nature.com/news

BUSINESS WATCH

The US solar-power industry got a boost in October with six plant approvals from the Department of the Interior — including one last week for a 1 GW solar thermal project in southern California, currently the largest planned globally. Nat Bullard, an analyst at Bloomberg New Energy Finance in San Francisco, California, expects that about 6 GW of solar thermal capacity should be installed worldwide by 2013, although this is still dwarfed by booming photovoltaic capacity (see chart).

SOLAR PROSPECTS

At current growth rates, Germany will dominate the world's photovoltaic capacity in the coming decade.

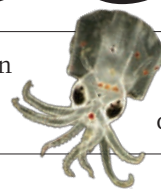


NEWS IN FOCUS

ASTROPHYSICS Stars in a spin may power gamma-ray bursts **p.15**

POLICY An end to aid for ex-Soviet weapons researchers **p.16**

DNA The master craftsmen who fashioned the 'synthetic cell' **p.22**



OCEAN SCIENCE Climate threat to a billion-dollar fishery **p.26**



NASA

Schemes to reflect sunlight away from Earth by injecting sulphur into the atmosphere have been called a threat to global biodiversity.

CLIMATE CHANGE

Geoengineering faces ban

Moratorium on schemes to reduce global warming clashes with reports urging more research.

BY JEFF TOLLEFSON

A last-ditch remedy for an ailing planet, or a reckless scheme that could be a greater threat to life on Earth than the problem it aims to solve? Opinions are sharply divided on geoengineering — potential massive interventions in the global climate system, intended to forestall the worst effects of climate change.

Last week, participants in the international Convention on Biological Diversity (CBD) made their views clear at a meeting in Nagoya, Japan. They included in their agreement to protect biodiversity (see page 14) a moratorium on geoengineering “until there is an adequate scientific basis on which to justify such activities and appropriate consideration of the associated risks”. The moratorium, expected to be in force by 2012, isn’t legally binding, and given the preliminary nature of studies in the area it is unlikely to affect

researchers in the near future. But some scientists fear that the CBD’s stance will sow confusion and delay at a time when governments and research groups are exploring how geoengineering might feasibly be undertaken if global warming accelerates disastrously.

The CBD agreement coincides with the release of a pair of reports on geoengineering, including a US congressional analysis, published on 29 October, that calls for research across the federal government. In his foreword to the report, Bart Gordon (Democrat, Tennessee), the outgoing chairman of the House Committee on Science and Technology, highlights the dangers of stifling research and calls for a “rigorous and exhaustive examination” of geoengineering strategies.

➔ **NATURE.COM**
For more on the case for geoengineering research, see:
go.nature.com/6Is5wx

“If climate change is one of the greatest

long-term threats to biological diversity and human welfare,” says Gordon, “then failing to understand all of our options is also a threat.” His report singles out the US National Nanotechnology Initiative — a programme that incorporates research at 13 federal agencies — as a possible model for coordinating research.

The Nagoya agreement grants an exception for smaller studies conducted in a “controlled setting”, but only if they are thoroughly assessed and “justified by the need to gather specific scientific data”. Ken Caldeira, a geochemist who studies geoengineering at the Carnegie Institution for Science in Stanford, California, finds the agreement’s language vague and confusing. “What does ‘specific’ mean? Who is to determine the necessity of the data? How do I demonstrate a need to do anything?” he asks. Caldeira is also concerned that the agreement does not distinguish between controversial geoengineering technologies intended to ▶

► block out the sun and less problematic techniques, such as removing carbon dioxide from the atmosphere.

The CBD set a precedent for such language in 2008, when it called for a halt to ocean fertilization activities but allowed small-scale research projects to continue. That ban caused the German government to suspend an ocean fertilization experiment, dubbed LOHAFEX, in the Southern Ocean in early 2009, although the project eventually went forward. Pat Mooney, executive director of ETC Group, an environmental activist organization based in Ottawa, Canada, that lobbied for the geoengineering ban, says the issue was advanced by representatives from all parts of the globe, including Norway, Switzerland, Bolivia and the Philippines. "It won't stop small experiments, but governments will think twice before they allow anything on a larger scale," Mooney says.

David Keith, a geoengineering researcher at the University of Calgary in Canada, says he agrees with the basic message that small-scale research should move forward, but large-scale technology deployment should be put on hold. Keith commissioned a separate report, also released on 29 October, that explores in detail the costs of injecting sulphur into the upper atmosphere to deflect sunlight. The report suggests that a small fleet of specially designed aircraft could inject 1 million tonnes of sulphur into the stratosphere for just US\$1 billion–\$2 billion a year. That would offset more than half of the global warming so far caused by anthropogenic greenhouse gases, and could be scaled up. "It is certainly the first time anybody has got serious about this," says Keith, who paid for the study through a fund that he manages with Caldeira, and that is provided by Microsoft founder Bill Gates. Keith stresses that the goal is to get a grasp of what a geoengineering programme might actually look like, so that scientists can properly focus their investigations into impacts and consequences.

The Intergovernmental Panel on Climate Change (IPCC) is encouraging scientists to expand their geoengineering work, and is planning a meeting in June next year to discuss the scientific basis for geoengineering, its costs, impacts and side effects, and how to treat the issue in the next IPCC assessment. Ottmar Edenhofer, chief economist of the Potsdam Institute for Climate Impact Research in Germany and co-chairman of the IPCC's working group on climate-change mitigation, says that commitments to reducing greenhouse-gas emissions will probably not be enough to meet widely discussed climate goals, such as limiting global warming to 1.5–2°C over the next century.

"Geoengineering is one option, and it should be included in a portfolio of other options," says Edenhofer. ■

BIODIVERSITY

World gets 2020 vision for conservation

With a raft of agreements emerging from Nagoya, the next step is finding the cash to move beyond the blueprint.

BY ANJALI NAYAR IN NAGOYA, JAPAN

The world has a bold new plan to stem the loss of biodiversity, and share its benefits fairly. But the vision has not yet secured enough funding to turn it into reality.

Delegates to the Convention on Biological Diversity (CBD) summit in Nagoya, Japan, last week agreed to targets that would, if achieved, offer much greater protection for natural resources over the coming decade. During the next two years, countries say they will put into place strategies to boost the size of protected areas and restore lost habitats, based on 20 headline targets (for examples, see 'Biodiversity goals').

Although these 'Aichi targets' have broadly been welcomed as a step in the right direction by conservationists, "we still believe that much higher targets are necessary to maintain the full range of critical ecosystem services essential for human well-being," says Russ Mittermeier, president of Conservation International, an advocacy group based in Arlington, Virginia.

When the CBD came into force in 1993, it articulated three main aspirations: to conserve biodiversity; to use natural resources sustainably; and to share the commercial fruits of

genetic resources, such as medicinal plants and crop strains. But signatories admitted last year that they had failed to meet their 2002 goal of reducing biodiversity decline by 2010. They have also spent years arguing over the best way to regulate the exploitation of the world's flora and fauna.

Towards the end of the Nagoya talks, it looked as though discussion on that issue — known as access and benefit sharing (ABS) — would once again result in deadlock. But hours after the conference had officially ended, negotiators emerged with the Nagoya Protocol, a subsidiary to the main convention that lays out ground rules for ABS. The protocol will come into force in signatory nations once it has been ratified by at least 50 parties.

It outlines mechanisms to ensure that developing countries with valuable genetic resources are recompensed by businesses and governments who develop products from indigenous plants and animals. The protocol focuses on products developed after the treaty takes effect, but it urges governments to pay developing countries — in cash, or through technology-transfer agreements — even if the crucial genetic material was obtained decades ago. The United Nations Environment Programme, which administers the CBD, says that one possible mechanism for reimbursing developing countries would involve putting a proportion of profits into a special fund that could be tapped to support conservation efforts or build scientific capacity.

The protocol commits countries to each set up their own national agencies to manage the system, overseen by an international 'clearing house' that will also set out detailed codes of conduct and help draw up contracts between parties. "I won't say it's a miracle that we achieved this agreement, but it is surely historic," says Mittermeier.

Japan has already promised US\$2 billion to help achieve the Nagoya targets, with smaller sums pledged by other CBD signatories, including France, Norway and the European Union. But governments have agreed to defer a detailed funding plan until the next convention summit in New Delhi, India, in 2012. ■

BIODIVERSITY GOALS

Key 2020 targets

- Cut the rate of loss of natural habitats, including forests, by at least half
- Increase terrestrial nature reserves from 13% to 17% of the world's land area
- Increase marine and coastal nature reserves from 1% to 10% of the world's seas
- Restore at least 15% of the areas where biodiversity is classed as 'degraded'
- Safeguard at least 75% of threatened plant species in collections

➔ **NATURE.COM**

See *Nature's*

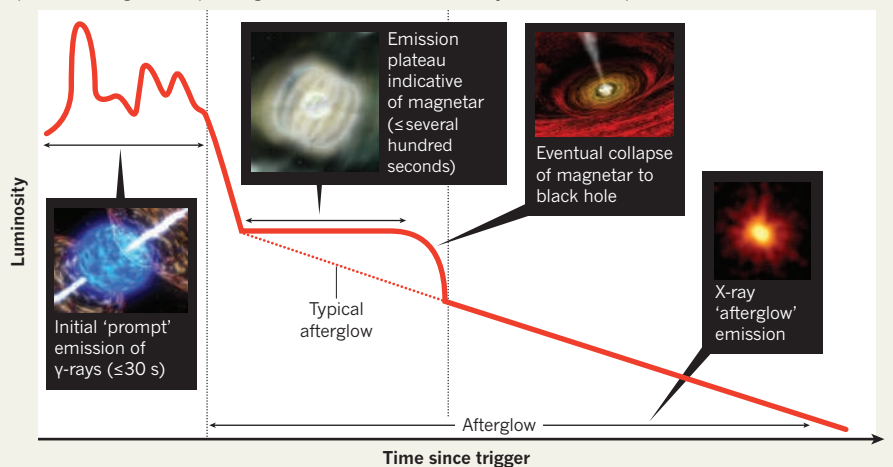
'Biodiversity in

crisis' issue at:

go.nature.com/dpclvc

PROFILE OF A MAGNETAR

Normally, the X-ray afterglow of a γ -ray burst fades rapidly, but the breakneck speed of a magnetar's spin flings out surface matter and delays the final collapse.



ASTROPHYSICS

Cosmic blasts hint at inner magnetars

A twist in the plot for the Universe's most powerful explosions suggests a detour en route to forming a black hole.

BY EUGENIE SAMUEL REICH

Powerful, mysterious and brief, γ -ray bursts are the flash bulbs of the cosmos. The brightest bursts outshine a million galaxies, then fade within minutes or hours. Only an extraordinarily energetic and localized event could generate such a brilliant flash. Most γ -ray bursts (GRBs) are thought to be triggered by the collapse or merger of stars to form black holes.

Until now, the short timescales and considerable variation among GRBs have made it hard to understand the physical processes at work. But recent observations from NASA's Fermi Gamma-ray Space Telescope, launched in 2008, and the Swift satellite, launched in 2004 and still going strong, have allowed astronomers to unpick the details of the explosions. Each mission has now detected more than 500 bursts, and results presented this week at the Gamma Ray Bursts 2010 Conference in Annapolis, Maryland, have added a twist to the standard fireball-to-black-hole plot.

The new findings point to the formation of magnetars, highly magnetized and rapidly spinning neutron stars, at the heart of GRBs. Although several magnetars have been observed within the Milky Way, their role as possible causes of GRB emissions has been largely ignored, says Neil Gehrels, an

experimental physicist who works with data from Swift at the NASA Goddard Space Flight Center in Greenbelt, Maryland. "Now it's a new hot topic," he says.

A GRB consists of an initial high-energy γ -ray and X-ray signal, called the prompt emission, followed by an afterglow of X-rays. 'Long' GRBs, which last for more than two seconds, are thought to be triggered when a massive star runs out of fuel and collapses. Jets of material rushing out of the collapsed object generate searchlight beams of radiation, which are visible from Earth if the jet is pointing in the right direction.

The supporting evidence for magnetars comes at the end of the prompt emission. In most cases, this fades rapidly. However, in 2007, Paul O'Brien at the University of Leicester, UK, and his colleagues reported on a GRB seen by Swift in which the X-ray emission held steady for hundreds of seconds before beginning its decline¹. Such a signal had been predicted by theorists² as the signature of a magnetar that forms within the GRB maelstrom and, for a while, spins fast enough to resist shrinking into a black hole (see 'Profile of a magnetar').

Brian Metzger at Princeton University in New Jersey, who is working on the theory of magnetars, says that the emission from the formation of a black hole would be expected to flicker as the amount of material falling in from the

collapsed star varies. By contrast, the emission from a magnetar is caused by material flung from its surface by a rotation rate that can be as much as one thousand per second, collimated by a magnetic field up to one quadrillion times that of Earth. Both the rotation rate and the magnetic field are expected to remain fairly constant for several hundred seconds before the magnetar spins down and finally collapses into a black hole.

One example was not enough to persuade everybody, says O'Brien, although "people were curious about it". But a sample of ten such cases³ makes the magnetar scenario much more plausible, says Metzger. "The black-hole model is still the more popular, but I think people consider both models as distinct possibilities," he says. Not all GRBs show the resulting plateau in X-ray emission, but they might still involve the formation of magnetars whose signature is swamped by the X-rays produced when the jets interact with gas surrounding the original star, says O'Brien.

Dissection of GRBs isn't just throwing up possible magnetars. At this week's meeting, Sylvain Guiriec at the University of Alabama in Huntsville presented the first detection of a long-predicted thermal component of the emission for GRB100724B, observed on 24 July 2010.

The thermal component is a measure of the black-body radiation that makes it possible to estimate the temperature of the explosion's photosphere, the apparent surface of the expanding fireball, which Guiriec says is about 44 million kelvin. The tricky part was distinguishing the thermal signature from radiation emitted by the jets. Guiriec spent two years surveying objects observed by the Fermi telescope before finding one for which the separation happened to be easy. He analysed another 40 rather more ambiguous bursts in the same way. Guiriec says this is the first time that astrophysicists can claim to have related a portion of the prompt emission to the physical model of the burst. "We were scared we'd be scooped," he says.

The next step will be to subtract the thermal emission from that of the entire GRB to help to explain the physical mechanisms that underlie the rest of the signal. Guiriec's first attempt to do so suggests that the standard picture is missing some important magnetic effects.

Gehrels says that future work is likely to focus on flashes from the demise of massive stars at the edge of the visible Universe, allowing astronomers to glean direct information about what the first stars to form were like. Both the European Space Agency's Cosmic Vision programme and NASA's Explorer programme, with respective deadlines of December 2010 and February 2011, are expecting proposals for GRB missions. ■

1. Troja, E. *et al. Astrophys. J.* **665**, 599–607 (2007).

2. Zhang, B. & Mészáros, P. *Astrophys. J.* **552**, L35–L38 (2001).

3. Lyons, N. *et al. Mon. Not. R. Astron. Soc.* **402**, 705–712 (2010).



Former chemical-weapons scientists at Russia's Shikhan complex received Western support.

WEAPONS RESEARCH

Curtain falls on collaborative work

End of the line for international funding agency that brought former Soviet weapons scientists in from the cold.

BY GEOFF BRUMFIEL

An international organization established to foster collaborations between Western researchers and weapons scientists of the former Soviet Union is set to close, *Nature* has learned. The Moscow-based International Science and Technology Center (ISTC) is now discussing its plans with the Russian government, which is increasingly irritated by the foreign handouts that the centre channels to the nation's weapons researchers. But some experts worry that the move — just months after President Dmitry Medvedev announced that Russia would withdraw its participation in the centre — will also sever a link between the country's weapons scientists and the rest of the world.

The centre was set up in 1992, just after the collapse of the Soviet Union. It was a dark time for thousands of scientists who had worked in the sprawling Soviet weapons complex, recalls James Toevs, an independent consultant and former US nuclear-weapons researcher based in Santa Fe, New Mexico. Shops were empty, and scientists went for months without pay. Toevs, who worked in Russia in the early 1990s, met nuclear-weapons designers who were forced to grow vegetables and do menial labour, such as repainting the local football stadium, to survive.

The United States, Japan, the European Union and the Russian Federation agreed to establish the ISTC to help alleviate the harsh conditions that scientists faced — and prevent weapons experts from taking their potentially dangerous knowledge elsewhere.

The centre organized international meetings

and disbursed generous grants to weapons designers who were willing to collaborate with counterparts in other countries. By the end of the 1990s, former Soviet satellite states were also benefiting — they, like Russia, did not have to pay into the scheme. From 1994 to 2009, the ISTC gathered more international partners and spent about US\$837 million on projects and meetings involving some 73,000 scientists (see 'Welfare for weapons researchers').

Not everyone was a fan of the programmes. "Some good science was done, but in the grand scheme of things it was ineffective," says Konstantin Severinov, a Russian-born virologist at Rutgers University in Piscataway, New Jersey, who worked on ISTC grants in Russia and in the former Soviet state of Georgia. Weapons scientists often performed low-quality research, and Severinov says that he saw cases in which extra scientists were added to projects unnecessarily to secure additional funding.

Adriaan van der Meer, ISTC executive director, concedes that in the early years, "we were probably more of a social programme than a research programme". But as time went on, the centre raised its scientific standards and even tried to help commercialize the work of researchers it supported. It also helped them find their way in the wider scientific world. "We got an opportunity to participate in international symposiums and to see how foreign scientists arranged their work," says Oleg Nagornov, a mathematician at the National Research Nuclear University in Moscow. Nagornov says that ISTC collaborations helped him to publish his first articles in Western journals, and he has maintained professional ties with some of the researchers he met.

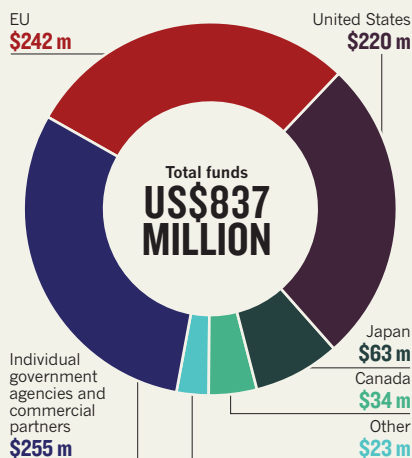
As Russia's economy boomed in the past decade, the government began to reinvest in its weapons complex. It also became increasingly resentful of the centre's role as a foreign funder of research within its labs, according to multiple sources familiar with the organization. Russia's irritation led the European Union to stop funding new ISTC grants on Russian soil this year. Meanwhile, the United States has begun to shift its non-proliferation funding towards hot spots in Asia and the Middle East — regions where the threat seems more imminent.

The ISTC plans to shut down its programmes gradually over the next three or four years. But van de Meer hopes that it can be replaced with a new organization in which Russia is an equal partner. There is still a need for collaborative work on non-proliferation science, and a new generation of researchers must be engaged so that they're fully aware of the potential pitfalls of their dual-use knowledge, he says.

Andreas Persbo, the executive director of the Verification Research, Training and Information Centre, a London-based arms-control organization, agrees that a line of dialogue between Russian weapons scientists and the West is vital. He fears that "without the ISTC, that will gently decline". ■

WELFARE FOR WEAPONS RESEARCHERS

From 1994 to 2009, the United States and the European Union both poured millions into the former Soviet Union's weapons labs.



SOURCE: ISTC

ARCHIVES

Researchers launch hunt for endangered data

Global effort will catalogue information languishing in drawers and basements.

BY LINDA NORDLING

Around the world, key scientific data are at risk of being lost, either because they are held on fragile or obsolete media or because they may be destroyed by researchers who are unaware of their value. Now a team of scientists is planning to scour museums and research institutes to draw up a global inventory of threatened data. Launched on 29 October, shortly after the biennial conference of the Committee on Data for Science and Technology in Stellenbosch, South Africa, the project aims to publish the inventory online in 2012.

The effort should enable future rescue programmes to target the most valuable and threatened material, and it should also alert researchers to resources of which they were unaware. Climate-change studies, for example, require data series on temperature and rainfall reaching back further than digital records. Some scientists are having to leaf through old ships' logs for clues to past weather patterns.

It is not only data in old books or photographs that are at risk. Digital information collected between 1950 and 1980 is also threatened, because it is stored on outdated media often subject to deterioration, such as magnetic tape and floppy disks, making it increasingly difficult to access and retrieve. Developing countries host some of the most valuable data — from land use to disease statistics — and the most threatened, for example by local conflict and inadequate or patchy storage. The task group hopes to catalogue such data by simply asking scientists, by e-mail and telephone calls, what records are lying in their vaults.

"I think the CODATA initiative is a very worthy one," says Sam Pepler, curation manager at the British Atmospheric Data Centre in Chilton. "A lot of our data heritage is locked up in an analogue form. Long-term,

accurate records are not an invention of the digital age, they have been kept for hundreds of years and the people collecting the data expected us to use them."

The inventory is the brainchild of Elizabeth Griffin, an astronomer at the Herzberg Institute of Astrophysics in Victoria, Canada, and William Anderson, an information specialist at the University of Texas at Austin and an associate editor of the *Data Science Journal*. Both will serve on the task group with ten other scientists.

"As things stand, researchers are only vaguely

aware of the existence of non-digital repositories of data in their own disciplines," says Griffin, who has relied on non-digital data in her own recent work. From old glass plates recording astronomical spectra, she has been able to study historic ozone levels in Earth's atmosphere. "I had to go from observatory to observatory with a suitcase, rummaging around these plates and books," she says. She estimates that some 3 million glass plates storing information about the night sky are gathering dust in observatories worldwide. Each is an



Millions of astronomical plates are just waiting to be rescued.

irreproducible window on the past, she says.

Some disciplines already have projects to recover old data. The Global Oceanographic Data Archaeology and Rescue Project, for example, has collected more than 3.5 million sea-temperature profiles — some dating back more than 100 years — since the effort began in 1993. Most fields are not so well organized, however, and it is unlikely that all the at-risk data that the task group identifies can be saved for posterity. As data sets start to be measured in petabytes (10^{15} bytes), scientific institutes are struggling to manage the vast streams of new data generated by projects, without adding to their burden with legacy data.

But without knowing where the old data are, says Anderson, it is impossible to make rational choices about what to keep. "I think everybody knows that stuff is going to be lost," he says. "But it would be nice to be on top of it." ■

➔ **NATURE.COM**
For more on data
curation see:
go.nature.com/vnklyt

'Biosimilar' drugs poised to penetrate market

Draft regulations will pave the way for copycat antibodies and other large molecules.

BY HEIDI LEDFORD

When is a copy good enough to be treated as the real thing? In recent days, regulators on both sides of the Atlantic have been grappling with the question as they try to develop guidelines that will expand the development of 'biosimilar' drugs: copycat versions of complex biological drugs, such as antibodies and other therapeutic proteins.

Biosimilars differ from generic drugs because their active ingredients are huge molecules with intricate structures. Such molecules are nearly impossible to replicate in every detail — even in the hands of the original manufacturer, minute variations in production yield slight differences. Unlike the relatively simple construction of a small-molecule drug, making a biosimilar is more like placing a complicated family recipe in the hands of a new chef. The overall result may be roughly the same, but it is not exactly how mother used to make it — and it may not precisely match the safety and therapeutic effects of the original.

Biosimilar medicines already have a foothold in Europe, but last week regulators there at the European Medicines Agency (EMA) hammered out draft guidelines that could drastically expand the market for these compounds. And in Washington DC, the US Food and Drug Administration (FDA), which in March received the authority to approve biosimilars as part of President Barack Obama's health-care reforms, is holding a public meeting this week to gather opinions on how it should evaluate such drugs. The new health-care act defines biosimilars as 'highly similar' to the original product; the FDA must now decide what that means, and how much extra testing will



Growth medium: Sandoz hopes to benefit from biosimilars.

be required of a biosimilar before it can be marketed.

"This legislation was anticipated for so many years," says Deborah Shelton, a partner at the law firm Sheppard Mullin Richter & Hampton in Washington DC. "Now it's here and we're still in a black hole."

Meanwhile, some companies see biosimilars as a low-risk way to bolster dwindling drug pipelines. Biological drugs are expensive, and even with a 20–30% reduction in the original price, biosimilars can still pull in a huge profit. "Not a day goes by when you don't read a press release saying some company is getting into biosimilars," says Michael Malecki, head of the biosimilars group at Decision Resources, a market-research firm based in Burlington, Massachusetts. But if regulators require extensive

testing, biosimilars could falter. For example, in 2006, the Croatian drug maker Pliva decided to abandon its copy of the best-selling anaemia drug erythropoietin after learning that the EMA would require more clinical trials than the company had anticipated.

Europe's framework for approving biosimilars was established in 2004, and copies of three drugs have already hit the European market. Some observers expect the FDA guidelines to be broadly similar to the European ones, which generally require clinical tests of biosimilars, but the extent and nature of the tests depends on the class of drugs being copied. A relatively simple and familiar molecule such as insulin, for example, may require less testing than a complex protein carrying several chemical modifications.

Yet biosimilars have had a slow start in Europe, says Huub Schellekens, who studies pharmaceutical development at Utrecht University in the Netherlands.

The three drugs copied in Europe — human growth hormone, erythropoietin and granulocyte colony-stimulating factor, another treatment for anaemia — were selected because they command a huge market, their patents had expired, and they are relatively simple to make. But these drugs were already embroiled in a price war with second-generation longer-acting versions. Furthermore, the inability to tap the US drug market may ultimately have stifled industry enthusiasm. North America accounted for 39.8% of world pharmaceutical sales in 2009 compared with 30.6% for Europe.

At a meeting on 27–28 October in London, regulators at the EMA worked to finalize draft recommendations for biosimilar versions of the biggest moneymakers of the biological world: highly specific

SANDOZ

MORE ONLINE

US MIDTERM ELECTIONS

Nature's special examines out what the future holds for science.
www.nature.com/midterm2010

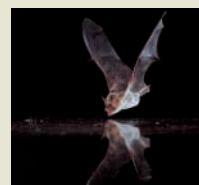


FOLLOW NATURE'S NEWS

- On Twitter twitter.com/naturenews
- On FriendFeed friendfeed.com/naturenews
- Receive daily or weekly e-mail alerts go.nature.com/jdlhcz
- Subscribe to the RSS feed go.nature.com/i2npj1

NATURE VIDEO

Stunning slow-motion footage shows how bats find water. go.nature.com/mznvym



TOP FIVE BEST-SELLING MONOCLONAL ANTIBODIES IN 2009

Once these biological drugs go off patent, they may inspire a wave of biosimilar imitations.

Name	Brand name	Target	Primary indication	2009 sales (US\$)	Estimated patent expiry
Infliximab	Remicade	TNF α antagonist	Rheumatoid arthritis	5.98 billion	2011 (US) 2014 (5 EU countries, Japan, global)
Bevacizumab	Avastin	VEGF	Colorectal cancer	5.73 billion	2019 (US) 2018 (EU) 2020 (Japan)
Rituximab	Rituxan/MabThera	CD20	Non-Hodgkin's lymphoma	5.61 billion	2016 (global)
Adalimumab	Humira	TNF α antagonist	Rheumatoid arthritis	5.49 billion	2016 (US) 2018 (5 EU countries, Japan, global)
Trastuzumab	Herceptin	HER2/neu receptor	Breast cancer	4.85 billion	2015 (US) 2014 (EU, global) 2016 (Japan)

antibodies called monoclonal antibodies. In 2009, monoclonal antibodies brought in US\$36.4 billion in sales worldwide. "These are blockbusters," says Falk Ehmann, the scientific secretariat for the Working Party for Biosimilar Medicinal Products at the EMA. "They are the hot molecules." The prospect of copying them — and selling the copies on the lucrative US market — could finally animate European biosimilars efforts, says Schellekens. The guidelines will be presented to the EMA's Committee for Medicinal Products for Human Use next week and, if accepted, will then be posted for public comment.

The first biosimilar monoclonal antibody is likely to be a copy of the cancer drug rituximab. Initially developed by Biogen Idec in Cambridge, Massachusetts, rituximab is expected to be one of the first monoclonal antibody

therapies to go off patent, which Malecki says it may do in Europe as early as 2013. It is also one of the best-sellers: in 2009 the drug earned about \$5.61 billion worldwide (see "Top five best-selling monoclonal antibodies in 2009"). In May, the Israeli generics maker Teva Pharmaceuticals announced that its biosimilar version of rituximab is ready for clinical trials, which are now expected to conclude in August 2011.

All biosimilars will probably require some clinical trials in the United States, says Mark McCamish, head of global biopharmaceutical development at Sandoz, a generics company based in Holzkirchen, Germany, but the FDA should expect a lively debate about how extensive those clinical trials must be.

On one side of this debate will be companies such as Denmark's Novo Nordisk, which derives at least 90% of its income from

biologics and has no biosimilar programme. Jim Shehan, the company's general counsel for North America, urges stringent clinical trials that pit a biosimilar head-to-head with the product it aims to replace.

Meanwhile, McCamish and others with a vested interest in marketing biosimilars — Sandoz already sells three in Europe and is developing about a dozen more — will argue that the FDA should use the variation in the original drug as a guide. Biologics are so complex that minor manufacturing changes often change the properties of the drug.

"There's going to be a range of how much a biosimilar company would have to do," says McCamish. "If the biosimilar falls within the goalposts of the originator's own changes, then you should have very abbreviated clinical testing." ■

Brazil's climate modellers are set to go global

Supercomputer will drive model to analyse effects of wildfire on world climate.

BY JEFF TOLLEFSON

Climate modellers in Brazil are getting ready to unleash a silicon god.

Named Tupã, after a native South American god of thunder, a Cray XT6 supercomputer is gradually coming to life at Brazil's Centre for Weather Forecast and Climate Studies in Cachoeira Paulista, northeast of São Paulo. By late November, when the machine is expected to be fully operational, it will have a peak processing speed of more than 244 teraflops, ranking it among the top 25 most powerful computers in the world. So great is Tupã's thirst for electricity that it will only run at 20% capacity until a new power source comes online in February 2011.

Working with their counterparts at the Hadley Centre in Exeter, UK, Brazilian modellers will use Tupã to apply their expertise on the release of black carbon and other aerosols from Amazonian fires to the global climate-modelling arena. "We want to bring knowledge of the Amazon into our model, and that will then be an original contribution," says Paulo Nobre, a senior modeller with the National Institute for Space Research (INPE).

Brazilian science minister Sergio Rezende proposed the initiative three years ago, as a strategic investment intended to nurture a relatively small climate-modelling team and help bolster Brazilian climate science on the international stage. Tupã builds on several decades of effort to develop a weather- and climate-modelling capacity; in time, the supercomputer could help to earn Brazil a place in the small club of nations that contributes global climate-modelling expertise to the Intergovernmental Panel on Climate Change (IPCC). China has paved the way among developing countries, but Brazil would be the first country in the Southern Hemisphere, apart from Australia, to develop such a capacity.

But the overriding goal is to develop domestic

modelling expertise that will help Brazilian scientists answer questions about national climate impacts and adaptations, says Henrique de Brito Cruz, scientific director of the São Paulo Research Foundation, which partnered with the federal Ministry of Science to fund the US\$30-million project. "What we want is for them to be able to develop and perfect their own models so that they can focus on questions and regions that are especially relevant for Brazil," says Brito Cruz. "To study

vegetation and climate, says Betts, and fire plays an important role.

Tupã will come online too late to participate fully in all the modelling scheduled for the IPCC's fifth assessment. Instead, says Nobre, the team will work with the Hadley Centre on century-scale climate simulations, while conducting regional modelling exercises that explore global warming's potential impacts throughout Brazil in more depth. Although hosted at the INPE, Tupã will serve

as a national supercomputer. The INPE is setting up a committee to analyse research proposals and allocate computer time; researchers will use the machine to assess a broad array of weather and climate issues.

Of particular interest is the potential impact of climate change on the Amazon region. The Hadley Centre's simulations have found that a warming of surface waters in the northern equatorial Atlantic Ocean could shift precipitation northwards, resulting in a significant dieback of the rainforest later this century. That is consistent with a major Amazon drought in 2005 and a second one that is under way at present.

Brazilian climate modellers have also begun collaborating with colleagues in South Africa and India and plan to host a new Earth system modelling workshop for scientists from all three countries next summer.

Francois Engelbrecht, an atmospheric modeller with South Africa's Council for Scientific and Industrial Research in Pretoria, says that Brazil's programme has helped inspire similar efforts to build up a global climate-modelling capacity in his country.

"It's a very good example of what can be achieved in a smaller but very focused modelling group," Engelbrecht says. ■



The supercomputer Tupã aims to take the world by storm.

regional issues, you need a global model."

Rather than starting from scratch, the Brazilian approach has been to piggyback on the Hadley Centre's Global Environmental Model (HadGEM2_ES) while adding new features that reflect the complex interactions between the rainforest and the atmosphere above it. The resulting model could be the first to incorporate a detailed treatment of aerosol emissions from fire, enabling UK and Brazilian scientists to probe how the climate system responds to hot plumes carrying black carbon and other chemical compounds produced by fires in the Amazon and around the world.

"It is a tremendously exciting step forward," says Richard Betts, who heads climate-impact modelling at the Hadley Centre. The IPCC's fifth assessment includes a systematic analysis of various land-use scenarios that will allow scientists to explore the interactions between

CORRECTION

In the News Feature 'The century of the city' (*Nature* **467**, 900–901; 2010), the population of Dhaka in Bangladesh in 2025 is expected to be 18.7 million, not 18.7 billion.

A. P. TAVARES/INPE

➔ **NATURE.COM**
For more on climate
modelling and the
Hadley Centre see:
go.nature.com/nlfkgx

DNA's master craftsmen

Behind the walls of the J. Craig Venter Institute, Ham Smith and Clyde Hutchison quietly worked to bring a synthetic cell to life.

BY ROBERTA KWOK

Sitting in adjoining offices on the second floor of the J. Craig Venter Institute (JCVI) in San Diego, California, Ham Smith and Clyde Hutchison carry on a fragmented conversation through their open doors.

"Clyde, did you do your timesheet?" says Smith. "It's due in 12 minutes."

Hutchison pauses: "No, it was due three hours ago."

They work at their computers for a minute in silence. Both are dressed in green short-sleeved shirts, tan trousers and black shoes, although they swear the wardrobe coordination is an accident. "The market finished flat," says Smith.

The dialogue is almost constant between the two men — friends and collaborators who are rarely seen apart. They are the veteran DNA craftsmen at the JCVI, a non-profit research organization founded by the provocative genome scientist who gave it his name. Originally from academia, the pair has taken advantage of Venter's resources and unparalleled salesmanship to pursue ambitious projects in synthetic biology, research that Venter says could enable faster vaccine production and the development of organisms that churn out precursors to fuel.

In May, the JCVI team announced that it had built a 1-million-base-pair genome — the longest working piece of chemically synthesized DNA yet assembled — and used it to restart a bacterial cell¹. Although some scientists disagree on whether the resulting micro-organism, called 'Synthia' in the popular press, is indeed 'synthetic' — the synthesized genome sequence was cribbed from a related bacterial species rather than being built to a novel design — few deny the technical skill demonstrated by such work. "The ability to synthesize and put together so many nucleotides without a mistake really requires guys on the level of Smith and Hutchison," says David Botstein, a geneticist at Princeton University in New Jersey who has worked with Smith. "I don't think many other people could have done it."

The two are acclaimed for their pioneering work. Smith shared a Nobel Prize in Physiology or Medicine in 1978 for his discovery of a restriction enzyme² — a protein that cuts DNA at specific sites. In the 1970s, Hutchison helped

to determine the first full sequence of a DNA molecule and co-developed site-directed mutagenesis³, a technique that enables researchers to make targeted changes to DNA sequences. The methods Smith and Hutchison helped to develop underpin much of the work done today in molecular genetics. "They're both regarded as scientist's scientists," says David Baltimore, a molecular biologist at the California Institute of Technology in Pasadena. "They've both done enormously important work in basic science."

Now 79 and 71, respectively, Smith and Hutchison have become intellectual partners and close friends. Smith, who looms over Hutchison, is slouchy, often sporting a half smile somewhere between amusement and embarrassment; Hutchison is more deliberate, precise in movement and dry in delivery. They are still pushing the possibilities of their field and, stints at the desk aside, they try to spend

WE DIDN'T KNOW IT WOULD BE POSSIBLE TILL IT WAS DONE.

about half their working hours at the bench — something that helps them maintain "a real perspective on what is possible," says Hutchison. This comes in handy when tasked with meeting the sometimes audacious goals set by their boss. "At one level, each of them operates like a septuagenarian postdoc," says John Glass, a synthetic biologist at the JCVI's other campus in Rockville, Maryland.

THREE'S COMPANY

When Smith and Venter first met in 1993, Venter already had a reputation. He was still five years away from challenging the public effort to sequence the human genome, but the US National Institutes of Health, where Venter worked until 1992, had filed patent applications on DNA fragments sequenced by Venter's team. The move didn't go down well with other scientists, nor did his tendency to antagonize scientific rivals. "He was commonly called an asshole," says Smith. But over drinks in a bar in Spain, the two discovered common ground: both started in medicine and had served in the navy. Venter invited Smith, then a professor at

Johns Hopkins University School of Medicine in Baltimore, Maryland, out for dinner with friends, and the group got drunk. "Almost the first instant that I actually met him, I liked him," says Smith.

Smith joined the scientific advisory council of the non-profit organization The Institute for Genomic Research (TIGR) in Gaithersburg, Maryland, which Venter had founded, and began collaborating scientifically with Venter, despite concerns of academic colleagues that the association might taint Smith's career. Their teams worked on determining the first genome sequence of an independently living organism, the bacterium *Haemophilus influenzae*⁴ — in which Smith and co-workers had originally discovered the HindII restriction enzyme. They chopped the 1.8-million-base-pair genome into pieces, sequenced the fragments and then assembled them computationally into a genome sequence. It was the first time the technique, called shotgun sequencing, had been used on such a large DNA molecule. Smith's excitement, colleagues say, outweighed practical considerations.

Jean-François Tomb, a former research associate at Johns Hopkins, recalls members of Smith's lab worrying about a lack of funding for the project, but Smith was only interested in the science. "He said, 'Look, you sequence the genome once and it's forever,'" says Tomb.

When the *H. influenzae* genome was nearly done, Venter wanted to sequence another right away. Smith said that they should try something small, and thought of Hutchison, then at the University of North Carolina at Chapel Hill. Hutchison was studying *Mycoplasma genitalium* — a bacterium thought to have the smallest genome, at half a million base pairs, of any free-living organism. Venter liked the idea. "He said, well, as soon as we leave lunch, why don't you call him up?" says Smith.

Hutchison agreed to help and sent Smith 10 micrograms of *M. genitalium* DNA. They finished the sequence in about two months⁵. "I was very pleased," says Hutchison. New techniques appealed to him, and he often pushed ▶

Clyde Hutchison (left) and Ham Smith (right) have forged a scientific partnership that allows them to go after questions few others would.

S. HUFFAKER/GETTY IMAGES



► his team to develop better methods. “From the beginning, Clyde was into high-throughput,” says Mike Conrad, a former postdoc in Hutchison’s University of North Carolina lab. “He liked stuff that was fast.”

When Hutchison took a sabbatical at TIGR in 1996, he, Smith and Venter began to discuss the idea of developing a cell with the minimum genome needed to survive. Hutchison was already investigating which genes *M. genitalium* could live without⁶, but he knew that deleting multiple genes simultaneously from this bacterium was technically difficult. The threesome speculated that they might need to synthesize whole candidate genomes and test them in recipient cells.

Hutchison again collaborated with Smith in 2003 at Venter’s latest non-profit institute, the Institute for Biological Energy Alternatives (IBEA) in Rockville. Their team synthesized the 5,000-base-pair genome of the bacteriophage ΦX174 (ref. 7). Hutchison had helped determine its sequence in the 1970s, and the small size made it convenient for trying out synthetic techniques. Smith and Hutchison had very different experimental styles, recalls team member Cindi Pfannkoch. “Clyde likes to plan everything,” she says, whereas Smith practices more casual ‘bathtub biochemistry’. In spite of this, the two got along. “They speak the same language,” she says.

TAMING THE CELL

In 2005, Hutchison started full time at the JCVI, which was formed by a merger of the IBEA and other Venter organizations. Leaving university life has disadvantages: “I can’t do whatever I want,” says Hutchison. “We’re not totally independent agents.” But the trio’s interest in big scientific challenges has kept them together. “I think all three of them are more about just doing the home-run experiment,” says Dan Gibson, a synthetic biologist at the JCVI in Rockville.

The synthetic-cell project picked up steam in 2005 as more researchers joined Venter’s synthetic-biology team, which eventually comprised some two dozen scientists. The project evolved into a two-pronged effort: one group focused on constructing a synthetic *M. genitalium* genome, while the other tried to transplant natural genomes into cells of different *Mycoplasma* strains and species (see ‘The path to a synthetic cell’).

Smith and Hutchison, who worked on genome construction, were circumspect about the chance of success. “We didn’t know it would be possible till it was done,” says Hutchison. Technical challenges loomed. The large DNA segments might break; the slow growth rate of *M. genitalium* limited the pace of progress; and — most crucially — rebooting a cell with a new genome had never been done before.

Although often seen just as the public face of the JCVI, Venter contributed to the science. While Smith and Hutchison worked out the

details, Venter made key strategic decisions. At first, the team tried assembling pieces of the *M. genitalium* genome from short DNA fragments rather than ordering longer, prefabricated, but more expensive segments from a DNA synthesis company. But progress was slow. “After we’d been working on this a couple of months, Craig comes into the lab and says, ‘Ham, how many pieces do you have put together so far?’” says Smith. “And I said, ‘Well, we haven’t got any of them yet.’ And he says, ‘All right, we’re going to order them.’”

I JUST WANT TO UNDERSTAND, THOROUGHLY, ONE CELL.

The genome-transplantation group was having no luck either. Carole Lartigue, a postdoc on the team, worked on the problem for two years “with nothing but failure,” says Glass, who oversaw the research. Lartigue finally got the first evidence of successful transplantation in late 2006, and the following year the team announced that it had managed to transplant a natural — not synthetic — genome from *Mycoplasma mycoides* into cells of the related species *Mycoplasma capricolum*, changing their identity⁸. Although the paper was announced to great fanfare, the mood at the JCVI wasn’t always so jubilant. A few months after publication, Smith came into the lab distraught. He feared that some *M. mycoides* cells, from which the donor genome was originally isolated, might have become patched up and revived when mixed with the recipient cells — an idea he called the ‘punctured tyre’ hypothesis. Resurrected cells could have been mistaken for transplants.

Hutchison, who is known for being meticulous about experimental controls, says he “thought the evidence was pretty good that

wasn’t the case”. But Smith’s doubts were not laid to rest until late in 2008, when the team started up *M. capricolum* cells with a natural *M. mycoides* genome that had first been inserted into and modified in yeast — a process that ‘purified’ the genome of any cellular remnants of its original host⁹. “That was absolute proof,” says Smith.

By now, Smith and Hutchison were working from the JCVI’s new lab in San Diego — a move prompted, according to Smith, by Maryland’s freezing winters. In 2003, Smith had slipped on the ice and broken his leg. A couple of years later, while he and Venter were walking through the sleet in Rockville, “I turned to Craig and said, ‘Why do we live here?’” says Smith. “He said, ‘That’s a good point.’” Venter set up a lab near the University of California, San Diego, his alma mater, and Smith and Hutchison moved west. “I came out because Ham was coming out,” says Hutchison.

By this time, JCVI researchers had also chemically synthesized the *M. genitalium* genome. Starting with segments of 5,000 to 7,000 base pairs made by DNA synthesis companies, the researchers connected them into progressively larger pieces of DNA, first *in vitro*, and then in yeast¹⁰. But they still couldn’t successfully transplant the *M. genitalium* genome, and the organism’s slow growth rate meant that they had to wait at least a month to see the results of an experiment. “We just didn’t know if it would ever work,” says molecular biologist Gwynedd Benders, a former team member. “It’s like, are we just going to keep hammering at this?”

As early as 2007, Venter had suggested switching the donor species: synthesizing the *M. mycoides* genome — which was roughly twice as large as *M. genitalium*’s — and transplanting it into *M. capricolum*. *Mycoplasma mycoides* grew faster, and the team had already managed to transplant the bacterium’s natural

THE PATH TO A SYNTHETIC CELL

1995: Genome sequence of the bacterium *Mycoplasma genitalium* is published.

1996: Smith, Hutchison and Venter begin discussing the idea of a minimal cell.

2003: Smith and Hutchison’s team synthesizes the genome of bacteriophage ΦX174.

Bacteriophage ΦX174
5,386 base pairs
assembled from
~260 pieces.

2005: Synthetic-cell project picks up momentum at the J. Craig Venter Institute (JCVI).

2006: JCVI researchers first successfully transplant a natural *Mycoplasma mycoides* genome into *Mycoplasma capricolum*.

2007: Assembly of chemically synthesized *M. genitalium* genome is completed.

2008: The team decides to try synthesizing the *M. mycoides* genome for transplant instead.

March 2010: A synthesized *M. mycoides* genome is successfully transplanted into *M. capricolum*.

Mycoplasma genitalium JCVI-1.0
582,970 base pairs
assembled from
>100 DNA
fragments.

Mycoplasma mycoides JCVI-syn1.0
1,077,947 base pairs assembled
from >1,000 DNA fragments.

genome. The researchers had been reluctant to switch because they didn't yet know whether the *M. mycoides* genome could be transplanted from yeast, where the synthesized genome would need to be assembled, into bacteria. But once they had successfully done this with the natural genome, Venter's idea seemed more attractive. In late 2008, Venter discussed a strategy for the *M. mycoides* genome synthesis with Smith, and Smith e-mailed Gibson, "Craig wants this done PDQ" — pretty damn quick. "It's that kind of thing where Craig is absolutely essential," says Smith. "He was pushing it. I would have dragged for months."

WAITING GAME

But synthesizing the *M. mycoides* genome presented new problems. Pieces of this genome above a certain size didn't replicate well in *Escherichia coli*, the bacterium used to amplify the DNA. Even determining the *M. mycoides* genome sequence proved difficult because of repetitive DNA sequences. "It just seemed to go on for months and months," says Benders.

After the *M. mycoides* genome was finally assembled, the team endured some suspense-filled weekends. Transplantation experiments were performed on Fridays, and Gibson checked for the blue bacterial colonies that would indicate success on Monday mornings. Week after week, no blue colonies appeared, and finally a single mutation was discovered in one of the synthetic DNA segments. This was corrected and transplantation tried again on Friday 26 March 2010. "That was a really, really long weekend," says Gibson.

On the Monday morning around 6 a.m. Gibson found a single blue colony and e-mailed Venter, Smith, Hutchison and Glass. Knowing tests were still needed to confirm the result, Gibson says he was "sweating" all day about the possibility of contamination, even as they celebrated with champagne. But Smith, who had told him to wake Venter that morning with the news, was more sanguine. Gibson says, "He just knew it was real."

Since the announcement, the team has fielded criticism for calling the resulting cell 'synthetic' when the genome was essentially a replica of a natural genome and required an existing recipient cell. Hutchison argues that 'synthetic' simply means 'chemically synthesized', not newly designed, and recipient cell contents are eventually replaced. "You'd like to design a genome from scratch," he says. "You'd like to put it into a cytoplasm that you built up from scratch. But we're trying to do something we can do." Although many synthetic-biology researchers are tweaking existing genetic elements, assembling them into new combinations, and inserting the 'circuits' into different organisms, few aspire to design entirely new genes.

On a July morning, Smith and Hutchison sit together in an auditorium at the JCVI's San

Diego building listening to a presentation by summer intern Nico Enriquez. Members of the synthetic-biology team, whose lab is down the corridor from Smith and Hutchison's offices, are scattered in the audience; researchers in Rockville watch the talk by videoconference. The synthetic-biology team is now attempting to develop the 'minimal cell': Gibson and his colleagues are building new versions of the synthetic *M. mycoides* genome with genes removed, then transplanting the edited genomes into recipient cells and monitoring colony size and growth. Enriquez presents methods proposed by Smith to assess cells' growth rates by measuring DNA and protein levels, and initial results look promising. "Seems like Ham's got something right again," says Enriquez.

At the end of his talk, Enriquez shows a picture of a pair of frolicking otters. "They kind of remind me of Ham and Clyde," he says to laughter in the audience. "The tall one is Ham, and the shorter one is Clyde — always together."

Once the researchers have a minimal genome, they hope to determine the function of every uncharacterized gene and build a computer model that predicts the cell's responses to genetic changes. But such a system isn't necessarily more informative than an organism with a larger genome, argues geneticist

ALL THREE OF THEM ARE MORE ABOUT JUST DOING THE HOME-RUN EXPERIMENT.

George Church at Harvard Medical School in Boston, Massachusetts. Although the cell will yield some scientific insights, he says, they will probably be specific to *Mycoplasma*. "As you delete these things, you'll end up with a cell that is weaker and weaker, less and less industrially useful, and less and less relevant to sophisticated organisms," he says.

Smith and Hutchison agree that *Mycoplasma* is unlikely to be used industrially as it is expensive to grow, but say that lessons learned from developing and studying this system might apply to organisms better suited for commercial purposes. Hutchison says a smaller system will be easier to understand, and he predicts that some gene deletions might actually make the cell grow faster. Smith notes that scientists initially couldn't see all the uses of restriction enzymes, which have proved essential for manipulating DNA. And in any case, he is largely driven by curiosity. "I just want to understand, thoroughly, one cell," he says. Venter, too, emphasizes that the minimal cell is meant to be a proof of concept. "It's truly basic science," he says.

JCVI researchers are investigating whether their genome-transplantation techniques can be extended to other, more complex, bacteria, such as cyanobacteria.

➔ NATURE.COM
For more on
synthetic biology:
go.nature.com/hamqjj

The team is also trying to make large-scale changes to cyanobacterial genomes, with the eventual goal of enabling production of valuable chemicals, says Glass. On the more commercial side, Venter and Smith's company, Synthetic Genomics, based in La Jolla, California, has filed patent applications on the JCVI's methods, and aims to engineer algae that produce hydrocarbons that can be converted into fuel. A separate company, Synthetic Genomics Vaccines, set up by the JCVI and Synthetic Genomics, is working with pharmaceutical company Novartis, based in Basel, Switzerland, on ways of making flu vaccine production more efficient.

Although other synthetic biologists are already using the JCVI's techniques to assemble pieces of DNA, many agree that it will be a while before anyone designs a whole genome from scratch. Right now, the quickest route to industrial application is unquestionably the modification of existing organisms, says Venter. But, he says, "the future will be designing and making whole new species".

It is now down to Smith and Hutchison to make Venter's ideas a reality — their working styles distinct, but complementary. At a lab meeting, Smith sketches a plan on a whiteboard for an experiment to detect very small changes in bacterial growth rate, which will be necessary to compare *M. mycoides* strains with different genome versions.

"You should do multiple dilutions each time," says Hutchison.

Smith hesitates. "I want it to be very simple, so..."

"Yeah, but maybe in working out how to do it, you need to do something that's not quite as simple," says Hutchison.

The two have no plans to retire yet. Most weekdays they take a walk together in the hills around their office, which overlook Interstate Highway 5, ruminating on ideas and keeping an eye out for rattlesnakes. "I think this is the pinnacle of my career," says Smith, who wants to keep working for at least another four or five years. Hutchison adds, "But maybe we'll do something else next." ■

Roberta Kwok is a freelance writer in the San Francisco Bay Area.

- Gibson, D. G. *et al. Science* **329**, 52–56 (2010).
- Smith, H. O. & Wilcox, K. W. *J. Mol. Biol.* **51**, 379–391 (1970).
- Hutchison, C. A. *et al. J. Biol. Chem.* **253**, 6551–6560 (1978).
- Fleischmann, R. D. *et al. Science* **269**, 496–512 (1995).
- Fraser, C. M. *et al. Science* **270**, 397–403 (1995).
- Hutchison, C. A. *et al. Science* **286**, 2165–2169 (1999).
- Smith, H. O., Hutchison, C. A., Pfannkuch, C. & Venter, J. C. *Proc. Natl Acad. Sci. USA* **100**, 15440–15445 (2003).
- Lartigue, C. *et al. Science* **317**, 632–638 (2007).
- Lartigue, C. *et al. Science* **325**, 1693–1696 (2009).
- Gibson, D. G. *et al. Science* **319**, 1215–1220 (2008).

The tiniest catch

BY WENDEE HOLT CAMP

Marine scientists are prowling the Bering Sea to learn how climate affects minute sea creatures and the lucrative fishery that depends on them.



Themisto libellula.

at 2 a.m. in the far north, marine scientist Alexei Pinchuk started his work as the summer sun dipped below the horizon. Using a winch, he lowered a 10-metre-long net — split like a giant pair of trousers — over the starboard side of the RV *Thomas G. Thompson*, which was cruising in the Bering Sea, west of Alaska. He dropped the net to the sea floor and then hauled it back up to examine his catch.

Pinchuk, a Russian working at the University of Alaska's Seward Marine Center, had harvested a whole zoo of minute animals, including larval fish, squid, crabs and octopuses, plus amphipods, copepods and krill.

"They're much more interesting than mammals," he says, wryly. These miniature zooplankton form critically important middle links in the food web, helping to sustain the rich life of the Bering Sea. The region provides half of the commercial seafood caught in the United States annually, including cod, sole, flounder, salmon and crabs. The tiny animals have also given Pinchuk and

his colleagues a breakthrough in understanding how climate fluctuations affect Alaska walleye pollock (*Theragra chalcogramma*) in the largest single-species fishery in North America. The Bering Sea yields one million tonnes of pollock each year, amounting to one billion dollars. But the pollock population has begun to fall and scientists are concerned that climate change could put further stress on the fishery.

For the past three years, Pinchuk has been studying this region as part of the six-year, US\$52-million Bering Sea Project, a collaborative effort between the US National Science Foundation (NSF) and the North Pacific Research Board. The project has funded more than 100 principal investigators, including oceanographers, biologists, geochemists, social scientists, economists and ecosystem modellers, as well as their technicians and students.

At its heart, the Bering Sea Project is an attempt to understand how climate change may affect the region's fisheries, by studying the ecosystem from top to bottom. Researchers hope this strategy will help fishery

managers to protect the pollock and other species as the Bering Sea warms.

"The Bering effort is one of the few times a truly multidisciplinary effort has come together to create an integrated model of such depth, largely from scratch," says Beth Fulton, an ecosystem modeller with the Australian Commonwealth Scientific and Industrial Research Organisation's Marine and Atmospheric Research division in Hobart, which helps to manage Australia's fisheries.

HOME AWAY FROM HOME

As he surveyed his haul of plankton, Pinchuk swirled the container, causing tiny crustaceans called copepods to emit a fluorescent blue glow. Although many of the researchers on board thrilled at the occasional sightings of whales and short-tailed albatrosses, Pinchuk remained single-minded in his plankton obsession. "I hate birds. Walruses are ugly," he jokes.

This summer Pinchuk spent several weeks on board the *Thompson*, which left port just after the last of the winter sea

ice melted in mid-June. Carrying 30 scientists, the ship cruised a zigzag trajectory across the eastern Bering Sea shelf, stopping every few hours to collect water, mud and samples of marine organisms (see 'Fishing for answers').

It will take some time for Pinchuk to sort his specimens. But he has already observed a distinct difference between the dominant zooplankton species during the past three years — the coldest on record — and the three previous decades, which culminated in five of the warmest years on record, from 2001 to 2005. An Arctic amphipod, *Themisto libellula*, that hasn't been seen here since the 1970s, is invading the Bering Sea now that temperatures have dropped. Pollock diets have preferentially shifted to these large, fatty invertebrates and to the big copepods and krill that thrive in cold years.

For researchers, one of the big questions about the Bering Sea

PHOTOS: T. SHAW

➔ **NATURE.COM**

Follow the author's blog postings from a Bering Sea Project cruise at: go.nature.com/omidii

food web is how year-to-year changes in ocean temperature influence plankton abundance, in turn affecting pollock and other commercial fish. A decade ago, oceanographer George Hunt of the University of Washington in Seattle developed an overarching model called the oscillating-control hypothesis. In simple terms, this suggests that the survival and growth of the youngest pollock fluctuate with shifts in the timing of the sea-ice melt in spring. Warm years help the cold-blooded zooplankton grow faster, and they provide ample food for the pollock born that year. In contrast, the hypothesis predicts that zooplankton and young pollock would not do so well in cold years.

"That was the first attempt to develop a broad conceptual framework for how climate might be impacting the Bering Sea," says Hunt. He approached the NSF for funds to test his hypothesis, but the agency instead asked Hunt to develop a broad programme for the Bering Sea, allowing others to examine his hypothesis in a much more comprehensive way. Hunt became chairman of a steering committee that planned the Bering Sea Project.

HOT AND COLD

As it turns out, zooplankton collected by Pinchuk and his colleagues during the project and in the warm years before it have both supported the oscillating-control hypothesis and turned it on its ear.

"The original hypothesis considered zooplankton as one group, and assumed they would all do better in warm years," says Pinchuk. New data, however, paint a more nuanced picture.

Small zooplankton species thrived in warmer conditions, as predicted, but larger, fatter species did not. The youngest pollock eat small copepods, but as the fish get bigger, they prefer energy-dense large copepods, amphipods and krill, says Pinchuk. As a result, the pollock initially fare well in warm years, but then start to falter for lack of nutritious food, and are unable to lay down enough fat to survive winter.

Warmer conditions can also hurt by revving up the metabolism

of the pollock, increasing the amount of food they need, says Franz Mueter, a fisheries biologist from the University of Alaska in Juneau who is collaborating with Pinchuk and others to revise the hypothesis.

As part of the Bering Sea Project, Mueter looked at pollock data going back to 1964 and found that survival of the younger fish does not improve in a linear way with temperature, as the oscillating-control hypothesis predicts. Instead, the relationship is a bell curve: they don't survive winter well in extreme warmth or in extreme cold. (Unusually chilly water during the past few cold years has forced the young pollock into relatively warmer regions, where they can more easily fall prey to older pollock.)

The results gleaned from work in the Bering Sea do not bode well for pollock, says Hunt. Although the sea has been colder than normal recently, temperatures have risen significantly since the 1980s and forecasts suggest that trend will continue. Juvenile

made up of state fisheries agencies, the general public and federal agencies. Each year, the North Pacific Fishery Management Council reviews data on the pollock population in the Bering Sea and recommends limits on how much the industry can catch.

The Bering Sea Project is providing information that could help to establish a much more sophisticated strategy — one that takes into account not just the size of the pollock population but also zooplankton numbers, predator data and many other factors to decide how much pollock should be caught each year.

The big question is whether this comprehensive strategy will be ready fast enough. The Bering Sea pollock fishery is regarded as one of the best-managed fisheries in the world, and the Marine Stewardship Council certified it as sustainable in 2005. But shortly after that, pollock stocks dropped dramatically, perhaps because of the cold snap. Last year, research-



A larval squid caught by researchers with the Bering Sea Project.

the conservation organization Oceana. He thinks that the Bering Sea Project could help in creating a better management system.

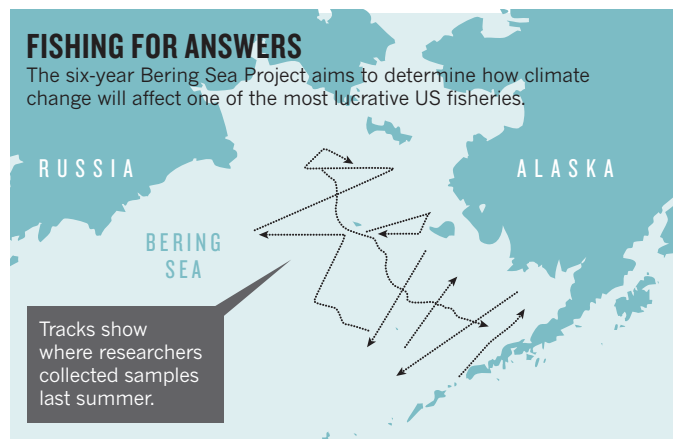
Not everyone agrees that the system is broken. Jim Ianelli is a scientist with the National Oceanic and Atmospheric Administration who runs the pollock-stock assessment models used to set current catch limits. "I think we're harvesting sustainably and responding appropriately by scaling catches according to trends in abundance," says Ianelli, who is also a principal investigator on the Bering Sea Project. He says that information gleaned from the project and other studies has already helped managers make more informed decisions.

On board the *Thompson* last summer, researchers hurried to collect as much data as possible in the short time available. After Pinchuk finished gathering his harvest, the ship headed off through steely grey waves towards a horizon shrouded in fog.

Pinchuk has been studying the northern seas for more than two decades, long enough to witness the changing climate firsthand. He has watched Alaska's glaciers melt and zooplankton populations shift in response to rising temperatures.

Arctic amphipods have shown up in his hauls, lured south by the cold snap of the past three years. But if temperatures resume their climb, as expected, Pinchuk suspects the amphipods will disappear once more. And with them, some of the pollock may also head north into Russian territory, in search of cooler waters. ■

Wendee Holtcamp is a freelance writer in Houston, Texas.



"Prospects for the pollock population are grim."

pollock suffered in the hot years of 2001–05 and if they do not fare any better as the water warms in the future, "prospects for the pollock population are grim," says Hunt.

Researchers are trying to give the fish a fighting chance by providing the foundation for a new way to manage the Bering Sea fishery. Like others in the United States, it is managed by a council

ers who conducted surveys in the area were shocked to find relatively few three-year-old fish.

That caused Greenpeace to move pollock to its 'red list' of seafood to avoid. Some researchers worried that the population was heading for a crash.

"It's quite possible to take too many fish for the ecosystem to function as it once did," says Jon Warrenchuk, ocean scientist with

COMMENT

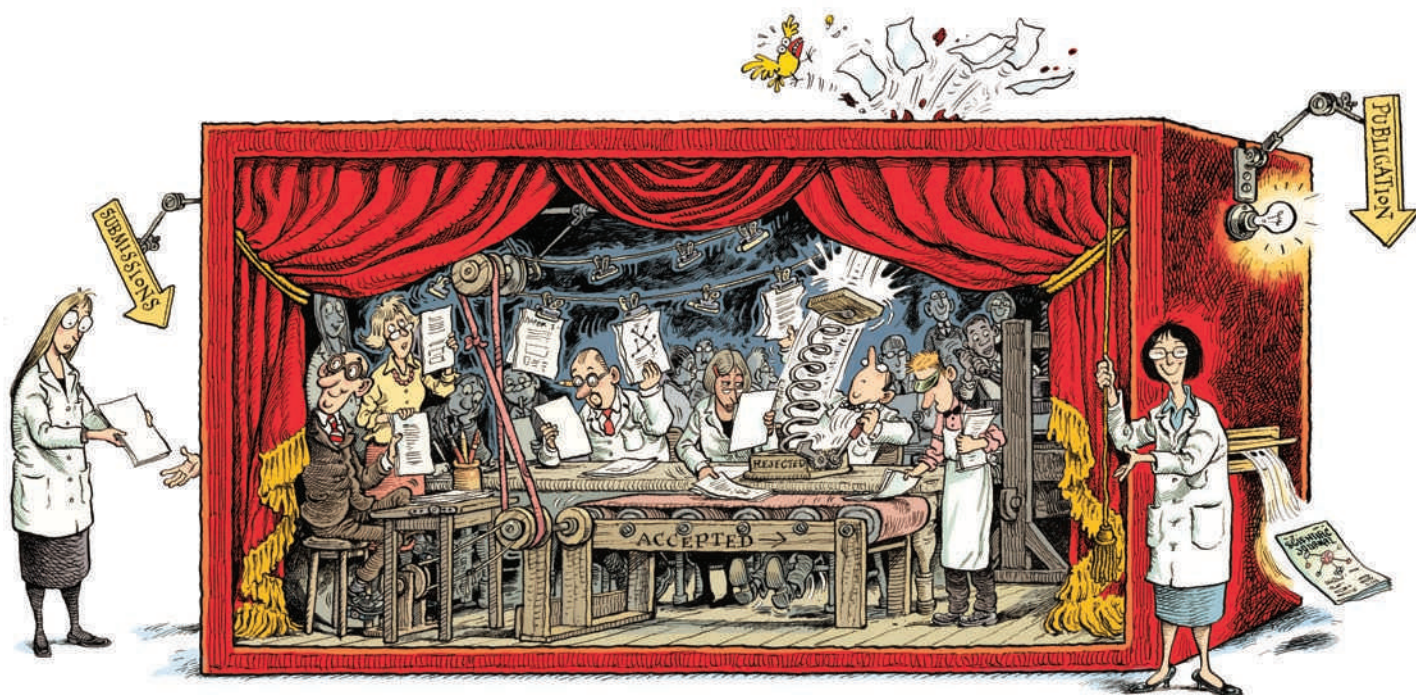
AUTOBIOGRAPHY Warts-and-all reflections on a career in social psychology **p.32**



DEVELOPMENT The long shadow of life in the womb **p.34**

FILM Science film festival judge muses on his misgivings about movies **p.35**

CODE Incentives needed to encourage programmers to publish. **p.37**



Transparency showcases strength of peer review

Bernd Pulverer reflects on his experience at *The EMBO Journal* of publishing referees' reports, authors' responses and editors' comments alongside papers, as other EMBO publications adopt the same policy.

Two years ago at *The EMBO Journal* we added transparency to peer review. We invited authors to allow inclusion of 'peer-review process files' alongside their published papers. Almost all have agreed. Now, more than 400 primary papers published in the journal showcase details of the editorial process: referee comments from every round of revision, editorial decision letters, the authors' response, as well as a detailed timeline of submission, decisions, revisions and publication^{1,2} (see go.nature.com/nbus3f for an example of an *EMBO J.* process file).

In our view, these augmented papers are testament to the fact that carefully administered peer review works — works well, in fact. We were initially concerned that some authors and referees might be discouraged from contributing to the journal and so, until now, have made the files relatively hard to find. But, given the positive response from the community, we are this month extending the policy to all four European Molecular Biology Organization (EMBO) scientific publications — *The EMBO Journal*, *EMBO Reports*, *Molecular Systems Biology* and *EMBO Molecular Medicine* — and making

the process files much more visible online.

The perennial concerns voiced about peer review and decisions made by professional editors — as opposed to part-time academic editors — stimulated us to think about how we might improve the process at EMBO. As a first step, we did a detailed annual analysis of where manuscripts rejected at our journal were eventually published, a summary of which we now publish annually (see go.nature.com/4y7fwf). This supported our sense that editorial decisions are generally informed and fair. For example, only 1% of manuscripts rejected in 2008 ended up in ▶

ILLUSTRATION BY DAVID PARKINS

► journals with an impact factor two points or more above that of *The EMBO Journal*; and only 9% have a citation rate higher than the average paper in the journal.

Our second thought was that a huge amount of effort goes into peer review — effort that remains largely invisible. Many an editor and referee will attest to how much the process can improve a published paper — painful as it may be to go through. Referees can be the best writers of published analyses of single papers, such as *Science's* Perspectives and *Nature's* News & Views. So why hide all their incisive, constructive comments, which can remain pertinent even after revision and publication?

An obvious solution was to publish our anonymous referee reports. It would showcase the quality and thoughtfulness of the majority of reports. And it would add interesting points about suggested further experiments, alternative interpretations and, sometimes, limitations.

Another appeal of this path was that peer review is rarely formally taught, yet so much depends on it. We hoped that the peer-review process files might serve as a teaching tool. Finally, a clear potential benefit was to fortify the peer-review process. Referees might feel compelled to take extra care when writing their report, as the report would be published, albeit anonymously.

It was immediately apparent that, for completeness, we'd have to post all referee reports on a paper, followed by the author response. In the spirit of transparency and accountability, and with the hope of addressing grumblings about professional editors, we decided to add editorial decision letters. We'd only correct simple typos in the reports, but we'd allow removal of data that were provided solely to address a referee's point, as they might be required for future publications.

IMPACT ASSESSMENT

The policy kicked off in January 2009 (ref. 1). We invite authors to opt out of the system at any stage, and referees are made aware at invitation that their comments will be posted in case of acceptance. In September this year we decided to discourage 'confidential comments for the editor' by referees, which are commonplace at many biological sciences journals². Legitimate confidential comments are allowed — for example, notes about bio-

security or conflicts of interest. But we want to move away from anything that gives rejected authors the sense that something went on

behind the scenes that led to their rejection.

At the time, *Biology Direct* and a number of BioMed Central journals already included published reports and author responses. Nevertheless, as with any change to a long-established system, there were significant risks. Would we discourage trusted referees? Would they fear that their identities might be revealed, and would they write less incisive or less critical reports as a result? Would authors resent the airing of — in the words of one referee — the 'dirty washing' leading up to acceptance? What if reports were rude or even defamatory? Would divergent

referee opinions, but we feel that the reader is well aware that journals invite a referee panel with complementary expertise and thus different vantage points. It hasn't happened yet, but if we felt a referee's report was too aggressive, we'd go back to the reviewer to suggest a rethink, noting the possible publication of the comment.

So does anyone actually notice the process files? The numbers show that the access rate is about one-tenth that of the main paper and that almost all peer-review process files have been viewed. Access to the files correlates with access to the whole article. The most

viewed files are those of the papers that most excited the editors and reviewers — not of controversial or borderline papers (see graph). We haven't been collecting data on how long readers spend looking at the files, but plan to. And now we've made the files much more visible and open access, we expect an uptick in access. Meanwhile, other journals, including the *European Journal of Cell Biology* have been taking note and are implementing similar enhancements.

One crucial limitation of the policy is, of course, that we do not release reports on manuscripts that end up being rejected. It goes without saying that these are often the more interesting cases to consider. However, a workable way to redress this shortcoming has eluded us. A partial solution we're pursuing instead follows the example of the Neuroscience Peer Review Consortium. This cooperative of 37 journals has agreed to share referee reports if an author desires. 'Review recycling' is in our view an important way to address a key bottleneck in the publishing process.

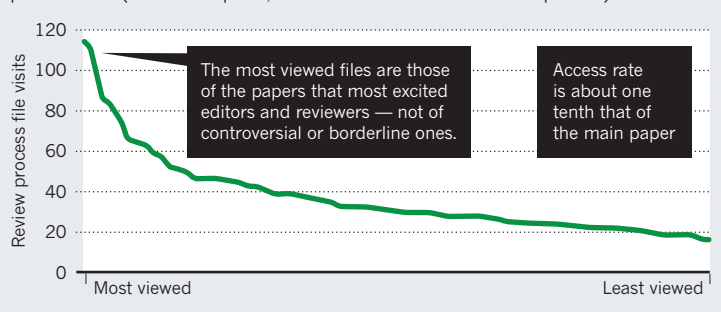
WHERE NEXT?

To mix metaphors, we feel we have pried open the black box of peer review with this initiative — and shown that it is not Pandora's box. Now, like many others interested in optimizing the scientific publishing and grant-review processes, we are considering several other enhancements to traditional 'single-blinded' peer review. We remain convinced that a high level of quality assessment is essential to filter and validate the increasingly vast and diverse literature.

Many journals now allow post-publication commenting, often curated and usually signed. Despite the ubiquity of social media, commenting on scientific reports has not yet reached a level at which it could give peer review a run for its money. Inspired by the physical and computational sciences, where pre-publication commenting is commonplace, *Nature* ran an interesting trial a few

PEER REVIEW PROCESS FILES ARE BEING NOTICED

Web access between April and July 2010 to the 100 most visited process files (referees' reports, editors' letters and authors' responses).



referee reports lend ammunition to those that believe the system is failing? More pragmatically, would producing the files increase our editorial costs significantly, and would this additional step slow down the publication process in a field in which every day can count?

The experience has been overwhelmingly positive. The number of submissions to the journal remains steady and just 5.3% of authors have opted out, few of them citing philosophical objections to the policy². The objectors cite a reluctance to add to the already excessive literature or a perception that an otherwise excellent piece of work can be marred by prominent comments on small mistakes or limitations.

The rate of acceptance of invitations to review a paper has remained the same, and very few invited referees decline explicitly because of the policy. In one case, a referee who had failed to read to the end of his invitation letter in the first round did decline to re-review the revised manuscript, but agreed to post the first set of comments. Nor have we seen a significant change in the quality of referees' reports or authors' responses — for better or for worse. Several referees have acknowledged that they spend more time on phrasing their reports now and this is certainly true for my own two-finger-typed decision letters! And we estimate that each file takes around 1½ hours for our administrators and editors to produce.

Many of the process files include divergent

years ago in which authors were invited to open up their manuscripts to pre-publication scrutiny during a 'traditional' peer-review process. Around 70 authors participated and the editors carefully compared the input received in this and a peer-review-alone approach. In no instance did commenting add significant value^{3,4} (see go.nature.com/n67mfk for the report).

Nevertheless, *Nature* opened all its published content for readers' online comments in March 2010. Comments, even on high-profile papers, remain sparse, however, even in journals such as *PLoS ONE* that specifically set out to supplement their assessment process with comments. Everyone is busy, and few may wish to risk outing themselves as critics without tangible benefit.

If peer review benefits from anonymity, why not also mask the author's identity ('double-blinded' review)? We remain interested in this possibility, but fail to see how to implement it without adding delays or requesting anonymized manuscripts for initial peer review (removal of author names does not suffice to anonymize a manuscript from one's peers). Conversely, why not add accountability by asking referees to sign their reports? The *British Medical Journal*, among others, has bravely pursued this path, and its editors claim that neither their referee pool

"Online comments, even on high-profile papers, remain sparse"

nor their reports have changed⁵. In our view, the stakes often remain too high for this in the competitive world of biological research. Can a rookie investigator really be expected to write a critical report on a manuscript submitted by an eminent colleague who may well review their next grant? Can an author who has been asked to revise a paper significantly be relied on not to persuade the referee to back down?

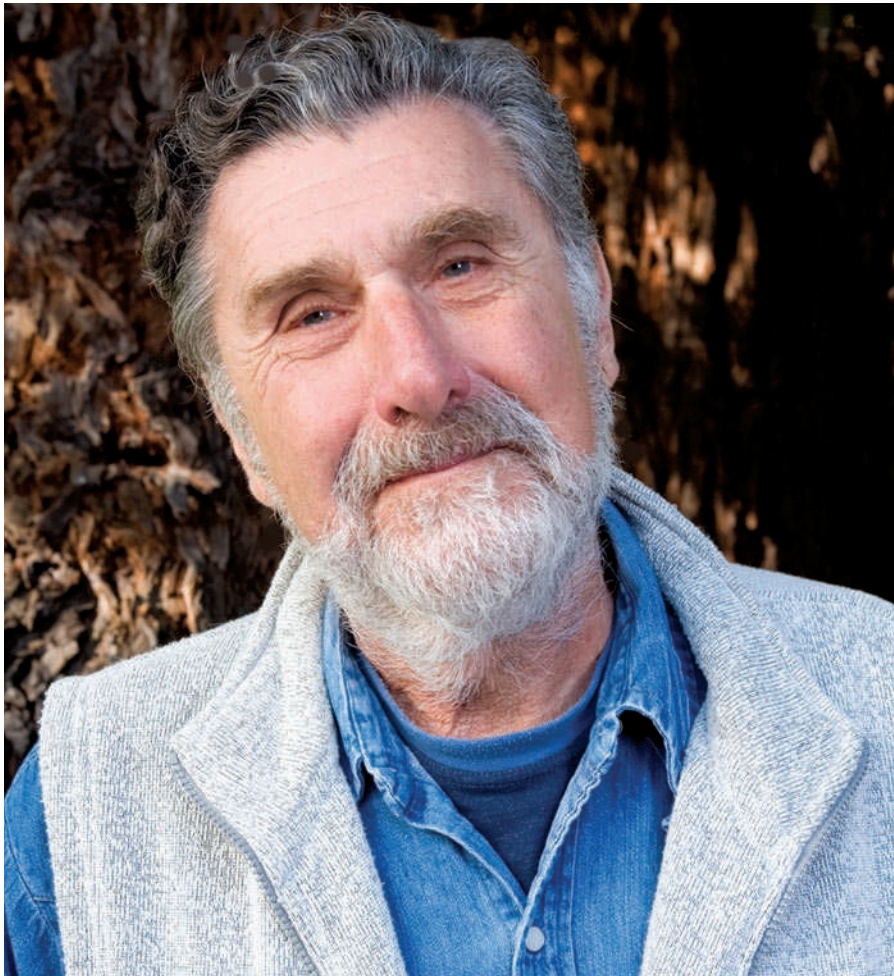
Last month we started to encourage referees to comment on each other's reports, where they feel this would aid the editorial decision. Comments are only expected in cases in which a referee has taken a particularly extreme line or made a mistake, or if a referee wants to underline an essential point made by a colleague that they had missed. In line with some other journals, we have also implemented another change: we now explicitly prompt reviewers to declare the common practice of delegating peer review to others in the lab. We request that reports are vetted by the invited referee and that co-referees are named. We regard this as an essential component of good mentorship.

Most successful scientists spend a good

fraction of their time reviewing papers. Yet, there is little tangible individual credit derived from the anonymous and voluntary contribution to this cornerstone of the research system. Thankfully, the remarkable culture of willingness to help colleagues and journals through peer review remains healthy, despite ever-increasing publication rates. Nevertheless, we are keenly pursuing means to allow funding agencies and tenure committees to take this essential activity into account, and we welcome suggestions and collaborations on this and other possible enhancements. Peer review is the most remarkable manifestation of a collaborative spirit of science and needs to be nurtured and fortified where necessary. ■

Bernd Pulverer is head of scientific publications at the European Molecular Biology Organization and chief editor of The EMBO Journal (published by Nature Publishing Group on behalf of EMBO), Meyerhofstrasse 1, D-69117 Heidelberg, Germany.
e-mail: bernd.pulverer@embo.org

1. Rørth, P. *EMBO J.* **28**, 1–3 (2009).
2. Pulverer, B. *EMBO J.* **29** (in the press).
3. Editorial *Nature* **444**, 971–972 (2006).
4. Editorial *Nature Cell Biol.* **9**, 1 (2007).
5. Smith, R. *Br. Med. J.* **318**, 4 (1999).



Elliot Aronson has studied the psychology of how people cope with conflicting beliefs and experiences.

PSYCHOLOGY

A social animal revealed

The inner strengths of psychologist Elliot Aronson are on display in his honest autobiography, finds **W. F. Bynum**.

Social psychologist Elliot Aronson might not be a household name, but his work is. As author of the landmark book *The Social Animal*, first published in 1972, he experimentally demonstrated the psychology behind Groucho Marx's famous line: "I refuse to belong to any club that would have me as a member." In Aronson's autobiography, *Not By Chance Alone*, he applies his scientific insight to his life and its dilemmas.

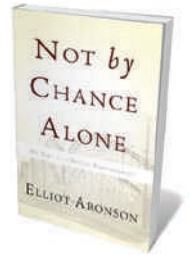
Aronson is best known among social psychologists for his work on cognitive dissonance, the mechanism by which

people cope with conflicting beliefs and experiences by minimizing discrepancies, thus reducing anxiety and tension. Building on the theory of one of his teachers, Leon Festinger, Aronson and his colleagues performed a series of classic experiments that extended its application. Most famously, they showed that the harder it is for people to become part of a group, the more they value membership — no matter how trivial the achievement needed to join.

Aronson has investigated many human traits, including hypocrisy, efficiency, attraction and sexuality. His experiments

are cleverly designed and he insists on rigorous standards. His co-authored chapter on experimental methods in the second edition of the *Handbook of Social Psychology* (Addison-Wesley, 1968) helped to make the field credible as a science.

Experimental social psychology is often criticized for relying on deceit and theatricality. Its test subjects must not know the point of the enterprise. Aronson believes that such experimental design is justified because it gives real insight into human behaviour. He describes how social psychologist Stanley Milgram discussed with him early plans for a controversial 'obedience to authority' experiment, in which subjects were told to inflict what they thought were increasing levels of pain on other participants by administering fake electric shocks. Milgram's results — that most subjects inflicted 'pain' on others when commanded — generated



Not by Chance Alone: My Life as a Social Psychologist
ELLIOT ARONSON
Basic Books: 2010.
304 pp. £15.99

ire from both the public and the scientific community. Aronson defends the tests as showing how ordinary citizens might have acted in the Nazi era.

Much of the autobiography is devoted to Aronson's career at several US universities, including Stanford University, California; the University of Minnesota in Minneapolis; the University of Texas at Austin and the University of California, Santa Cruz, where he is an emeritus professor today. He also taught at Harvard for a couple of years, from 1959 to 1961, but he liked the place little more than the place liked him. He valued the quality of the Harvard students, but felt that Ivy League elitism was not for him. As a Jewish academic, he was also realistic about the chances of getting tenure at a time when formal quotas limiting admission of ethnic minorities were still rife. He left Harvard before he had to.

The fact that he got there at all is a tribute to his personal qualities. His family was dysfunctional. His father coped badly with the loss of his small shop during the depression of the 1930s, and struggled with the dynamics of family life and the insecurities of being an inadequate breadwinner. Aronson writes eloquently about the ambivalence within him that his father's remoteness caused. He had to work hard not to become like his father — prone to storming out of the house when nagged by his wife, or to react with loud violence. The storming out, Aronson suggests, was probably calculated: his father had a gambling

addiction and his departure allowed him to spend the evening with his cronies.

Aronson's achievements are remarkable given the family's poverty, the anti-Semitism and bullying he experienced, and his early mediocre academic record. He describes his childhood in Revere, Massachusetts, with an eye for good stories and controlled recollection. His elder brother Jason — outgoing, clever and socially graceful — was everything that the younger Aronson was not, although his early hatred for him transmuted into affection and then, after Jason's premature death from cancer, into wistful family reflections. Aronson's younger sister is largely absent from his account.

Aronson keeps his youthful idealism closely under wraps, but his later career offered many opportunities for socially informed activity. While he was lecturing at the University of Texas in the 1960s, racial integration was of great significance and provoked high emotions. Aronson and his colleagues devised in the early 1970s a new method of classroom teaching for school children called the jigsaw technique. Small sets of students, containing children from different racial and ethnic backgrounds, and of different gender and ability, each cooperate in putting together part of the

"Aronson offers a revealing portrait both of himself and of social psychology in the United States during the past half-century."

answer to a larger question that the whole class has been posed. The strategy improved relations between children of different races and backgrounds, and enhanced the academic performance of the disadvantaged children.

After the system

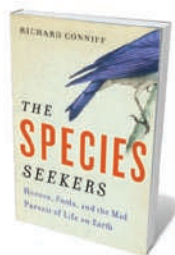
was eventually appreciated by educationalists, Aronson spent much time conducting classes for school teachers.

Today, macular degeneration has rendered Aronson almost totally blind, with the consequence that this book was an oral production. His greatest stated regret is that he can no longer distinguish his grandchildren by sight, although he did help one of them to learn to read.

Autobiographies are often carefully crafted, revealing only what the subject wishes to be known. Aronson, however, is happy to describe himself, warts and all. In doing so, he offers a revealing portrait both of himself and of social psychology in the United States during the past half-century. ■

W. F. Bynum is emeritus professor of the history of medicine at University College London, UK.
e-mail: w.bynum@ucl.ac.uk

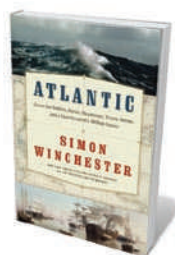
Books in brief



The Species Seekers: Heroes, Fools, and the Mad Pursuit of Life on Earth

Richard Conniff W. W. NORTON 464 pp. £19.99 (2010)

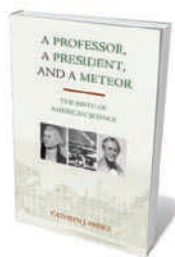
Celebrating the explorers who have scoured the planet for new forms of life, naturalist and writer Richard Conniff highlights how new discoveries generate scientific breakthroughs. He argues that the same spirit of adventurous curiosity that drove Swedish botanist Carl Linnaeus to derive his nomenclature of species and that spurred Charles Darwin to unearth evidence for evolution is in scientists today, as they hunt for creatures new to science in remote corners of Earth.



Atlantic: Great Sea Battles, Heroic Discoveries, Titanic Storms, and a Vast Ocean of a Million Stories

Simon Winchester HARPER 512 pp. \$27.99 (2010)

Simon Winchester's magisterial biography of the Atlantic Ocean portrays it both as a planetary feature and as a human stage — the "inland sea of Western civilization" along whose coasts humans have settled and traded for millennia. From its early navigation by the Norse and European sailors, to its setting for the slave trade and warfare, he recognizes the ocean's autonomous power. He also notes how its waters are changing as a result of pollution, overfishing and climate change.



A Professor, a President, and a Meteor: The Birth of American Science

Cathryn J. Prince PROMETHEUS BOOKS 304 pp. \$26 (2010)

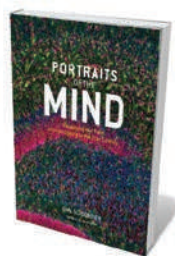
A meteor that lit up the sky in Weston, Connecticut, in 1807 sparked more than local consternation. It inspired Benjamin Silliman, a young chemistry professor at nearby Yale College, to analyse the meteorite's minerals. The controversy surrounding his hypothesis that it came from space triggered the rise of rigorous science in the United States. Cathryn Prince describes Silliman's life, his brushes with politics — including a spat with President Thomas Jefferson — and his other discoveries, among them the distillation of petroleum.



The Sorcerer's Apprentice: How Medical Imaging Is Changing Health Care

Bruce Hillman and Jeff Goldsmith OXFORD UNIV. PRESS 264 pp. £27.95 (2010)

Medical imaging has transformed health care, but its diagnostic value is balanced by its expense. Hillman and Goldsmith examine how scanning technologies such as magnetic resonance imaging and computed axial tomography are transforming radiology and yet putting pressure on budgets. They raise questions about economic incentives and the role of preventative medicine in promoting the uptake of scanning technology above other clinical necessities.



Portraits of the Mind: Visualizing the Brain from Antiquity to the 21st Century

Carl Schoonover ABRAMS 240 pp. \$35 (2010)

The human brain remains maddeningly complex. The intricacy of its structures is revealed by the striking pictures in this collection by neuroscientist Carl Schoonover. From medieval sketches and nineteenth-century drawings to state-of-the-art scans, he explains how our depictions of brain regions, neurons, axons and dendrites have developed throughout history. It is a fascinating visual insight into neuroscience.

The whole nine months

Michael Sargent examines the evidence that pre-birth experiences shape our lives.

The human uterus is not a tranquil nursery. At least one in 30 babies who make it through have health problems caused by genetic mutations and chromosomal misbehaviour. The fetus is also vulnerable to damaging external pressures that are less well understood. In *Origins*, journalist Annie Murphy Paul reviews the evidence that pre-birth experiences affect our health later in life.

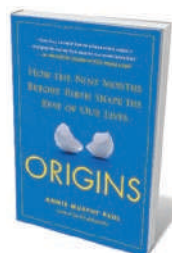
Paul recounts the mid-twentieth-century scandals, involving two drugs given to pregnant women, that made the biomedical community aware that fetuses could be sensitive to medications that are harmless to adults. Thalidomide, which was prescribed for morning sickness and approved for clinical use before it was common to rigorously test drugs for all eventualities, caused abnormal limb development. Diethylstilbestrol, given to prevent miscarriage, triggered a rare cancer in daughters. More fetus-damaging influences have been recognized since, including the neurotoxin methylmercury, alcohol, inhaled tobacco smoke, a number of recreational drugs and some viruses.

Investigations of the factors that affect human fetal development are difficult. But catastrophes can offer key insights. Studies of people conceived during the Dutch 'hunger winter' of 1944–45, at the end of the Second

World War, revealed how a fixed period of malnutrition affected the fetus. Many went on to develop type 2 diabetes and cardiovascular disease prematurely, and were more susceptible to schizophrenia.

Similar problems afflicted those who were conceived during the 1918 influenza pandemic. As well as having higher rates of diabetes and cardiovascular disease compared with a control group, this cohort did less well at school, were smaller as adults and were more likely to be poor. Investigations of nineteenth-century demographic data have also revealed that the season of conception affected the health and longevity of progeny, presumably because vitamin D synthesis triggered by sunshine and folate from fresh vegetables promoted healthy pregnancies.

In-depth studies of the fetal origins of ill health were initiated in the 1980s by David Barker of the University of Southampton, UK. His premise that the fetus is damaged by maternal undernutrition, resulting in



Origins: How the Nine Months Before Birth Shape the Rest of Our Lives

ANNIE MURPHY PAUL
Free Press/Hay House:
2010. 320 pp/256 pp.
\$26/£10.99

premature age-related diseases, now has widespread support after a controversial debut.

The fetus responds to poor nutrition by scaling down development. Unnecessary muscle growth is avoided, and excess resources are diverted into fat deposits that can be mobilized after birth when food is scarce. Animal experiments confirm that a diet deficient in protein or certain micronutrients is damaging to the fetus.

The resulting epigenetic modifications of crucial genes — changes that modify gene expression without altering the DNA sequence — can be transmitted to one or even two generations, Paul notes. So susceptibilities to diabetes and heart disease can be passed on from a historical episode of hardship. The optimal nutritional regime for human fetal development is frustratingly hard to formulate, although Paul's interviewees have strong views. Surveys do show, however, that later health benefits are conferred by a fish-rich diet and supplements of folic acid and vitamin D.

Paul is particularly fascinated by the evidence that maternal experiences of stress affect the fetus and influence the development of psychological characteristics. She recounts investigations of pregnant women and their progeny who experienced bereavements, Hurricane Katrina, earthquakes, famines and wars. Stress during pregnancy always involves constriction of blood vessels servicing the fetus and elevation of the stress hormone cortisol. The consequences include a risk of reduced birth weight, reduced cognitive and language skills, schizophrenia and premature age-related diseases.

NATURE.COM
The bioethics of clinical trials on pregnant women:
go.nature.com/WMephF

On a positive note, the author reports that tests show that improved prenatal care and nutrition can raise the IQ of offspring of disadvantaged women. Interestingly, Caesarean delivery is less stressful to the baby than a natural birth, and much less stressful than a birth involving forceps. The special character of the intrauterine environment is revealed in another way by the observation that autism is associated with exposure of the fetus to excess testosterone, although as yet there is no indication of why this should be.

Origins is an absorbing account of how we might be shaped by our early life in the womb — physically, psychologically, even temperamentally. ■

Michael Sargent is a developmental biologist at the National Institute for Medical Research, Mill Hill, London and author of *Biomedicine and the Human Condition*.
e-mail: msargent@nimr.mrc.ac.uk



Malnutrition and stress can affect the physical and psychological development of the fetus.



Animation *An Eyeful of Sound* envisions how people with audiovisual synaesthesia see the world, winning it a top spot at the 2010 Imagine Science film festival.

COMMUNICATION

Learning to love science films

Carl Zimmer comes around to the idea that science and movies can enjoy a happy union.

For two years now I've judged science films for the Imagine Science Film Festival, a week-long celebration of the genre that runs each October in New York City. It's a peculiar job, I confess, because I'm often underwhelmed by science on the screen. But the more I watch, the more hopeful I feel.

My association with the festival came out of an argument with Alexis Gambis, then a graduate student at Rockefeller University in New York. Over a cup of coffee, Gambis explained that when he wasn't slaving over fruitfly cells, he made movies. He also runs the Imagine Science Film Festival (of which *Nature* is a sponsor). I thought the festival a wonderful idea, but feared it might end in failure. I explained to Gambis my deep suspicions about whether science and film can ever enjoy a happy union. He responded with an invitation to become a judge.

It is odd that science and films have such a complicated relationship, given that films were born out of science. The invention of photography in the nineteenth century made it possible to capture a series of images and use them to create an illusion of movement. With

Imagine Science Film Festival
New York City,
15–22 October 2010.
go.nature.com/b8W3g6

the development of faster cameras, movies began to seduce the world. Each technical advance has brought change to the cinema, although not every change has resulted in artistic progress — witness Smell-O-Vision and *Piranha 3D*, for example.

For all that science and technology have delivered to Hollywood, scientists have received little back. Researchers portrayed in films bear scant resemblance to those in real labs. Some on-screen scientists are villains that must be destroyed by common-sense heroes. Others threaten nature with Promethean recklessness. Yet others are mavericks who find cures for cancer single-handedly in jungle tree-houses. And movies often distort science itself. Tornadoes, volcanoes, spaceships, viruses: all obey the laws of Hollywood, not the laws of Newton or Darwin.

Scientists have gnashed their popcorn buckets, wishing for something better. In 2008, the US National Academy of Sciences set up the Science and Entertainment

Exchange to bring scientists and Hollywood film-makers together for fruitful exchanges of ideas. Gambis's film festival serves a similar mission: its website announces that it "encourages a greater collaboration between scientists who dedicate their lives to studying the world we live in and film-makers who have the power to interpret and expose this knowledge, ultimately making science accessible and stimulating to a broader audience".

I'm not convinced such collaborations will achieve this goal often, or even whether they should. Exhibit A: Harrison Ford. Earlier this year, he played a biochemist searching for a cure for a genetic disorder in *Extraordinary Measures*, a fairly accurate story inspired by a book by reporter Geeta Anand. In 2008, Ford also played a scientist in *Indiana Jones and the Kingdom of the Crystal Skull*, a fairly accurate account of a comic-book fever dream. *Extraordinary Measures* earned a meagre US\$12 million, whereas *Indiana Jones and the Kingdom of the Crystal Skull* earned \$317 million. Hollywood is a place of business, not charity, and the marketplace

speaks clearly: people want their scientists with bullwhips, not pipettes.

Even if Hollywood directors dedicated themselves to achingly realistic biopics about Peter Medawar or Henri Poincaré, that might not be a good thing. Films should not be propaganda, bludgeoning us with messages about how valuable certain things or people are. At their best, films embody the conflicts in our societies, and give form to our inner lives in all their ragged glory. They can use real aspects of the world as their raw material, but holding them drearily to account is a mistake. *Citizen Kane* is about a newspaper editor; it would not have been a masterpiece if Orson Welles had kept asking himself “Does this make journalism accessible to a broader audience?”.

SAME OLD SCIENCE?

When I ask science-promoting friends about their favourite science-themed films, I hear the same titles: *Apollo 13*, an account of the near-fatal Moon mission; *GATTACA*, a dystopian look at genetics; and *Contact*, based on a novel about the search for alien life by Carl Sagan. *Apollo 13* is great fun and quite accurate, but is merely procedural. We learn more about the workings of a lunar capsule than the workings of an astronaut's soul. *GATTACA* divulges the shocking revelation that technology can be used for ill. *Contact* is buoyed by Jodie Foster's charismatic performance and scenes of astronomers at work. But it drowns in solemn television news reports and ponderous congressional testimony about how tiny and insignificant and precious we are. Film critic Anthony Lane summed up *Contact* perfectly in *The New Yorker*: “How a movie that began with the promise of such excitement can fritter itself away into these plaintive consolations, I have no idea.”

None of this can deny the pleasure of science-fiction films. But they're even more pleasurable when science is a mere



Independent science films, such as the cryonics comedy *Cold and Dry*, go beyond Hollywood clichés.

jumping-off point for wonderfully non-scientific monsters and emotionally resonant metaphors. It doesn't matter that the acid blood of *Alien* cutting through the floors of the spaceship *Nostromo* really ought to pierce the hull and kill everyone inside. It doesn't matter that the neurologist in *Eternal Sunshine of the Spotless Mind* could never wipe out every memory of an old girlfriend from a dingy little office somewhere in Queens. Both fictions are delicious.

The Imagine Science Film Festival is not burdened by Hollywood's bottom line, nor is it blessed with Hollywood's bank account. These short films — lasting from a few minutes to just under an hour — have shoestring budgets, yet they exceed what you might find at the multiplex. The best of them go beyond lazy clichés, such as the researcher obsessed with science to fill some inner psychological void or the evil corporation using science to control the planet. They find something original to say, and use the storytelling power of film to say it beautifully. One of my favourites is *Cold and Dry*, a Norwegian cryonics comedy [*sic*]. In 11 sweet, sly minutes, the film forces us to think about the grotesque lengths to which we can take science even with the best of intentions.

Even the documentaries push the boundaries. Last year's winner of the festival's Nature Scientific Merit Award, *Magnetic Movie*, is a meditation on the magnetism that suffuses our world (see go.nature.com/AGsPhy). On the soundtrack, University of California scientists describe the nature of magnetic field lines, while we watch an empty laboratory come to life with hoops, knots and

dancing tapestries. *Tagged*, one of this year's entries, follows the experiences of a man who implants himself with a radio-frequency identification chip to see what it is like to be part cyborg. And in another entry this year, *Meet the Elements*, the geek-idol band They Might Be Giants proves that chemistry can be the basis for an addictive music video (see go.nature.com/IsPbOt). I will be eternally grateful to them for causing my daughters to sing about the periodic table at the top of their lungs.

One of the biggest surprises for me is how effective a muse neuroscience can be. *Skhizein*, which won an honourable mention (see 'Science Oscars'), uses sophisticated computer graphics to transform the world as its hero descends into schizophrenia.

The 2010 Nature Scientific Merit Award went to *An Eyeful of Sound*, which joyously imagines the world as it is seen by people with synaesthesia. It makes those of us with ordinary brains jealous of those who can't help but mingle sight and sound. *Mario Borodine*, a short film about a young inventor, won the Nature People's Choice Award as the audience's favourite (see go.nature.com/XvEwkh).

As I finish another year's judging, these surprises give me hope that I will ultimately lose my argument with Gambis. I will be happy to admit defeat and head out to the cinema. ■

Carl Zimmer is a writer based in Connecticut. His most recent e-book is *Brain Cuttings: Fifteen Journeys Through the Mind*.
e-mail: carl@carlzimmer.com

SCIENCE OSCARS

And the winner is ...

Nature Scientific Merit Award

An Eyeful of Sound

Directed by Samantha Moore

Nature People's Choice Award

Mario Borodine

Directed by Emanuel Hoss-Desmarais

Honourable Mention

Skhizein

Directed by Jérémy Clapin

For more on the 2010 Imagine Science Film Festival: go.nature.com/XvEwkh

CORRESPONDENCE

Computer code: more credit needed

As a microbiologist who has published work based on in-house computer code, I empathize with the laments in your News Feature (*Nature* 467, 775–777; 2010). However, its emphasis on more training for scientists and students fails to consider a downstream issue — retaining the talent afterwards.

From personal experience, there is a common perception that computational analysis is just a tool to enable ‘real’ discoveries on the bench. Hence, there is little incentive to brush up on computer-related skills. A few graduate students from my laboratory have invested time and effort to become respectable programmers, despite their lack of a background in computer science, but they became disillusioned when they found that their hard work was considered to be of secondary importance to the science.

These talented individuals have since left for careers in finance, management and information technology, where the same programming know-how and problem-solving skills are highly appreciated. They now enjoy shorter hours, comparable pay and greater job security than a tenure-track assistant professor.

The corollary of Nick Barnes’s observation in World View that “most professional computer software isn’t very good” (*Nature* 467, 753; 2010) is that good programmers — regardless of their scientific background — are in demand everywhere. Teaching programming skills and best practices to scientists may indeed improve the quality of

software, but the software in question might not necessarily be for scientific research.

Herman Tse *The University of Hong Kong, Hong Kong, htse@hkucc.hku.hk*

Computer code: incentives needed

I don’t expect to see major changes to scientists’ habits about publishing their software without clear incentives for them to do so (*Nature* 467, 753; 2010).

Given that the present value system in science is based almost exclusively on the publication of journal articles, publishing software should help scientists to publish papers. Once journals require (or at least strongly encourage) authors to submit their code as supplementary material, scientists will learn the tools and techniques necessary to get their code into a publishable state, and their employers will grant permission to publish it.

As the reproducible-research movement has also been pointing out for a while, the quality of science is then likely to improve significantly.

Konrad Hinsén *Centre de Biophysique Moléculaire (CNRS), France, konrad.hinsén@cnrs-orleans.fr*

Computer code: a model journal

As a one-time developer of codes for climate models (these days, I’m more of a user), I would like to draw your readers’ attention to a peer-reviewed journal that explicitly tackles many of the issues Nick Barnes raises (*Nature* 467, 753; 2010).

The journal *Geoscientific Model Development* (www.geoscientific-model-development.net) strongly encourages publication of modelling codes alongside

detailed descriptions of models. It was founded because models are seldom subject to the same degree of scrutiny and peer review as the results they generate — even though modelling is central to research into climate science.

As Barnes so rightly states, model codes themselves are rarely or never published traditionally. Model descriptions often need scientific results for publication and so are pared to the minimum when they do appear. Issues of reproducibility, platform dependence, version proliferation and the real nitty-gritty of modelling all need to be addressed in the literature.

Dan Lunt *University of Bristol, UK, d.j.lunt@bristol.ac.uk*

Reviewer disclaims competing interest

Michael Mann and colleagues intimate that I had an undeclared conflict of interest in reviewing Roger Pielke Jr’s book *The Climate Fix* (*Nature* 467, 920; 2010), citing alleged connections between me and the George C. Marshall Institute in Arlington, Virginia, reputedly a centre of climate-change scepticism.

Contrary to the authors’ assertions, I did not write a report for the Marshall Institute. My relationship with the institute consists solely of participation in a 90-minute discussion in 2006 concerning how Congress deals with scientific issues (see go.nature.com/epxft1). Also taking part were a former member of the congressional Office of Technology Assessment and a current senior official of the US National Research Council of the National Academy of Sciences.

The audience in 2006, as I recall, consisted mainly of congressional staff members. I did not mention climate change and I received no payment for my participation.

Science journalists in Washington DC are often invited to address meetings on science-policy issues. I have taken part in similar discussions under the auspices of the Brookings Institution, the American Association for the Advancement of Science, the Association of American Universities, the Association of American Medical Colleges, and at dozens of universities.

Daniel S. Greenberg *Washington DC, USA, danielg523@aol.com*

More insights from Crick’s lost letters

As author of the 2009 book *Francis Crick: Hunter of Life’s Secrets*, I congratulate Alexander Gann and Jan Witkowski for their detective work on the previously lost correspondence between Francis Crick and Maurice Wilkins (*Nature* 467, 519–524; 2010). This material is invaluable for biographers and historians.

The rediscovered correspondence affirms my judgement — based on letters preserved with successive typescripts of Watson and Crick’s 1953 papers, now at the Wellcome Library in London — that Crick was very much in charge of events during this period. It also reveals the continuing strength of Wilkins’ friendship with Crick, which in 1959 found expression in his pleading with Crick not to abandon Cambridge in favour of the United States.

Robert Olby *University of Pittsburgh, USA, olbyr@comcast.net*

➔ **NATURE.COM**

Join in the code debate at:
go.nature.com/ed3hsl

CONTRIBUTIONS

Submissions to Correspondence should be sent to correspondence@nature.com.

Gravity's weight on unification

Much research in theoretical physics is inspired at least in part by the idea of unifying all of the fundamental forces of nature. An analysis of how gravity affects other forces at subnuclear scales has major implications for that idea. [SEE ARTICLE P.56](#)

GIOVANNI AMELINO-CAMELIA

One of the main goals of modern theoretical physics is to find a common framework that explains all of the fundamental forces of nature: gravity; the strong force that binds quarks into protons and neutrons; and the forces of the electroweak theory (which encompasses electromagnetism on macroscopic scales, and the 'weak' and hypercharge forces on subnuclear scales). On page 56 of this issue, Toms¹ reports results that could be significant for finding such a framework. The author describes a mathematical analysis of the behaviour of gravity at ultrashort distances and of how this most familiar of forces affects 'Abelian gauge' theories, which, notably, include the theory of electromagnetism and that of the hypercharge force.

Relevant to Toms's study¹ is the fact that coupling constants — parameters that are used to characterize the strength of forces — are actually not constant but vary with the distance scale of the physical process in which they are measured. This phenomenon is both well established experimentally and predicted by current theories, and is known as 'running coupling constants'.

The author's results have implications for what might be the next revolution in fundamental physics, a hint of which is found by extrapolating the experimentally established 'running' of the coupling constants down to distance scales much shorter than can be accessed experimentally. Within the successful standard model of particle physics, we can make such an extrapolation for the running of the coupling constant of the hypercharge force (α_Y) — which is described by an Abelian gauge theory of the type studied by Toms¹ — and for that of the strong (α_S) and weak (α_W) forces, which are described by non-Abelian gauge theories. The results are shown semi-quantitatively in Figure 1, together with a cruder estimate of the effective coupling constant of gravity (α_G).

Gravity is the only force for which we currently do not have a reliable 'fundamental description', which would be applicable at

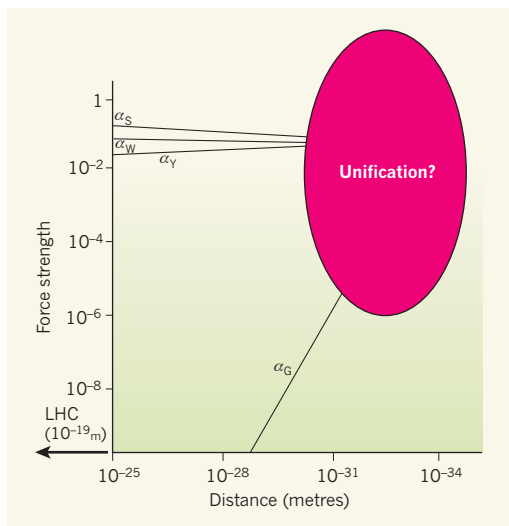


Figure 1 | Meeting at ultrashort distance scales?

Extrapolation of experimental values, measured at relatively large distances (not shown), for the strength of the strong force (α_S), weak force (α_W) and hypercharge force (α_Y) indicates that these forces might merge with gravity for processes characterized by distance scales that are at least 13 orders of magnitude smaller than those currently being explored with the Large Hadron Collider (LHC). The strength of gravity is characterized by α_G , and the strength of the electromagnetic force is not shown because at the distance scales displayed this force is replaced by the combination of the weak and hypercharge forces. Toms's analysis¹ investigates a piece of the puzzle that theorists must resolve to correctly describe this unification: the behaviour of gravity in the range between 10^{-32} and 10^{-35} metres.

subnuclear distances^{2,3}. But several indirect arguments invite us to give a preliminary description of its effective strength in terms of the ratio of Newton's constant to the square of the characteristic distance scale of the gravitational processes of interest.

It is striking that the coupling constants of the three non-gravitational forces converge with distance. And, perhaps even more remarkably, the effective coupling constant of gravity, which at currently accessible distance scales is much smaller than the others, also approaches the region where the other coupling constants converge. Within our current theoretical framework, this behaviour can only be described as an extremely fine-tuned numerical accident, but it clearly hints that some unknown law of nature is driving the convergence.

All of the most-studied proposals of modern theoretical physics are inspired, at least in part, by this possibility of unification. In particular, one of the motivations for conjecturing the existence of 'superpartners' of known particles⁴, which will be looked for with the Large Hadron Collider at CERN, Europe's premier high-energy physics laboratory near Geneva in Switzerland, is the role of these extra particles in the derivation of the running of the non-gravitational coupling constants. Including the roles of superpartners in the calculations leads to a particularly accurate convergence of the three non-gravitational coupling constants at a scale of about 10^{-32} metres.

Achieving unification of the non-gravitational forces at 10^{-32} metres has some intrinsic appeal, especially where the running of α_Y is concerned, as this would otherwise keep growing, possibly reaching values too high for the computational techniques of modern theoretical physics to manage. However, according to some theorists, by assuming such a stage of partial unification, with gravity left out because at 10^{-32} m it should still be too weak to matter, we fail to profit fully from the evidence summarized in Figure 1 — a single stage of unification of all forces, plausibly at distance scales a couple of orders of magnitude smaller than 10^{-32} m.

Toms's study concerns the weakest link on the road to unification: the behaviour of gravity at ultrashort distances, particularly in the range between 10^{-32} m and 10^{-35} m. In this regime, the lack of a fundamental, microscopic theory of gravity not only limits us to describing its strength crudely in terms of Newton's constant, but also prevents us from establishing its influence on the strength of the non-gravitational forces. To circumvent these limitations, Toms adopts a mathematical approach that exploits the fact that inconsistencies in our current description of gravity manifest themselves only at distances smaller than 10^{-35} m. Therefore, valuable insight should be obtained from computations arranged in such a way that only distances larger than 10^{-35} m are involved.

The author's analysis¹ suggests that the

effect of gravity on α_Y is such that, as soon as the strength of gravity becomes comparable to the strength of the other forces (at some point between 10^{-33} m and 10^{-35} m), α_Y quickly becomes vanishingly small, a behaviour known as asymptotic freedom. A mechanism for gravity-induced asymptotic freedom of coupling constants has been advocated⁵ by Robinson and Wilczek. However, it relies on a more preliminary analysis, which has been received with some scepticism. By providing evidence in favour of the Robinson–Wilczek mechanism, Toms removes one of the perceived advantages of having a stage of partial unification, because the unwanted, uncontrollable growth of α_Y is avoided. What's more, Toms's study strengthens another preliminary observation reported by Robinson and

Wilczek⁵ — that the effects of gravity on all three non-gravitational coupling constants combine to give a slower convergence of the non-gravitational forces, with the implication that the unification might occur at distances smaller than 10^{-32} m, at which it is easier to imagine gravity joining in.

Possibly even more important, Toms's findings provide encouragement for the idea that the role of gravity in the unification can be fruitfully investigated. And it might open the way to studies that compare different proposals for the sought-after fundamental description of gravity on the basis of their implications for unification⁶.

Perhaps we are one step closer to figuring out this amazing unification puzzle, and it is particularly exciting that this might rely on

modelling the effect of gravity on distances much smaller than those of planets and galaxies. ■

Giovanni Amelino-Camelia is in the Department of Physics, Università di Roma La Sapienza, and the Sezione Roma 1, INFN, Piazzale Moro 2, Rome 00185, Italy. e-mail: giovanni.amelino-camelia@roma1.infn.it

1. Toms, D. J. *Nature* **468**, 56–59 (2010).
2. Amelino-Camelia, G. *Nature* **408**, 661–664 (2000).
3. Carlip, S. *Rep. Prog. Phys.* **64**, 885–942 (2001).
4. Ellis, J. *Nature* **448**, 297–301 (2007).
5. Robinson, S. P. & Wilczek, F. *Phys. Rev. Lett.* **96**, 231601 (2006).
6. Calmet, X., Hsu, S. D. H. & Reeb, D. *Phys. Rev. D* **81**, 035007 (2010).

IMMUNOLOGY

Conditional stability of T cells

Data from several recent studies on the dynamics of regulatory T cells — which suppress excessive immune responses — do not add up. Collective analysis of the observations may reconcile the differences between them.

SHIMON SAKAGUCHI

Regulatory T (T_{reg}) cells — as their name suggests — are crucial for suppressing excessive immune responses that can harm the organism¹. They also establish tolerance to non-self antigens such as organ transplants. Deficiency or dysfunction of these cells in rodents and humans causes various T-cell-mediated autoimmune diseases, allergies and immune disorders such as inflammatory bowel disease. To understand how T_{reg} cells regulate immune responses, it is crucial to determine how they are produced and whether they are functionally stable under various conditions. A paper published in *Science* (Rubtsov *et al.*²) is the latest of several recent reports^{3–6} that address these questions. The papers find apparently conflicting results, suggesting that the T_{reg} -cell population can be both stable and dynamic.

T_{reg} cells specifically express the transcription factor Foxp3, and a cardinal feature of these cells is that, in the thymus, they are produced as a functionally mature T-cell population specializing in immune suppression¹. Some Foxp3-expressing (Foxp3⁺) T_{reg} cells can also differentiate from naive T cells, which do not express Foxp3, in the periphery — for example, in the mucosal layer of the small intestine, where immune responses to digested food or products of commensal microbes could be damaging

to the host and so must be suppressed.

Differentiation of naive T cells to either T_{reg} cells or effector T cells, such as helper T (T_H) cells, is elaborately controlled by immune mediators called cytokines^{1,7} (Fig. 1). *In vitro*, for instance, one cytokine, TGF- β , mediates differentiation of naive T cells to Foxp3⁺ T_{reg}

cells in response to antigenic stimulation, and another cytokine, IL-6, hampers this differentiation; together, TGF- β and IL-6 facilitate T_H cell differentiation into a type of mature T_H cell that secretes IL-17 (T_H17 cells) — a highly inflammatory cytokine. The cytokine IL-2 is indispensable for the maintenance of thymus-derived Foxp3⁺ T_{reg} cells and augments TGF- β -dependent T_{reg} -cell differentiation in the periphery; however, it inhibits differentiation of T_H17 cells.

Foxp3⁺ T_{reg} cells can also differentiate 'sideways', halting expression of Foxp3 and differentiating into effector T cells that secrete pro-inflammatory cytokines. This, however, may happen only under certain conditions, such as genetically determined reduced expression of Foxp3; antigen stimulation of T_{reg} cells *in vitro* in a T_H -cell-driven pro-inflammatory cytokine milieu; or *in vivo* transfer of T_{reg} cells to a T-cell-deficient environment, in which

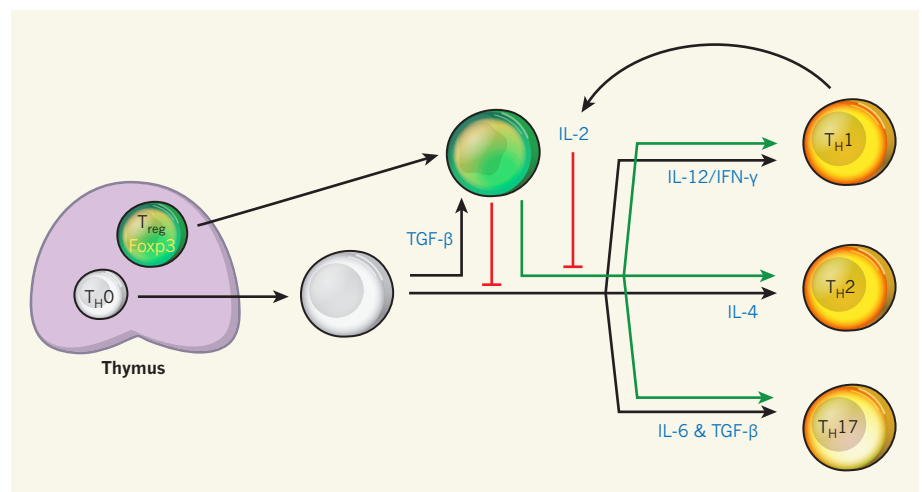


Figure 1 | T-cell differentiation. As well as Foxp3⁺ T_{reg} cells, which suppress excessive immune responses, the thymus produces naive T cells (T_H0 cells), which can differentiate into effector T cells, including T_H1 , T_H2 and T_H17 cells, following stimulation with antigens and in the presence of appropriate cytokines (blue). Although Rubtsov *et al.*² find that Foxp3⁺ T_{reg} cells are generally stable, under certain conditions at least some of these cells may stop expressing Foxp3 and convert into effector T_H cells. IL-2 produced by activated non- T_{reg} cells inhibits this conversion and is also required for the maintenance of Foxp3⁺ T_{reg} cells.

they undergo homeostatic proliferation^{3–5,8,9}. The conversion of T_{reg} cells to other types of effector T cell — if it occurs readily and frequently — could be harmful, because the resulting effector T cells, like the T_{reg} cells from which they are derived, are likely to recognize self antigens, and so may cause autoimmune disease.

Immunologists are on the quest to assess this possible danger. There are, however, conflicting results from experiments in mice that are engineered to express a reporter dye in their T cells whenever the cells express Foxp3; the dye-expressing T cells are then traced to determine whether they lose Foxp3 expression and become effector T cells. An earlier report⁶ showed that a significant fraction of Foxp3⁺ T_{reg} cells (about 10%) stop expressing Foxp3 and differentiate into T_H cells that secrete the pro-inflammatory cytokine IFN- γ and that mediate type 1 diabetes in a mouse model of autoimmune disease. Rubtsov *et al.*², however, show that T_{reg} cells are highly stable in terms of both Foxp3 expression and their suppressive function, with few cells converting into effector T cells even after exposure of the animals to an inflammatory cytokine milieu or to X-ray radiation, which causes a reduction in the level of circulating T cells.

Although apparently conflicting, these findings provide insights into how T-cell differentiation to and from T_{reg} cells depends on Foxp3 expression, on the mix of cytokines and on the presence of other T cells (Fig. 1). For one thing, a paucity of IL-2 clearly plays a crucial part in T_{reg} conversion: when Foxp3⁺ T_{reg} cells are transferred to T-cell-deficient mice, the co-transfer of T cells that do not express Foxp3 or infusion of IL-2 prevents conversion of the Foxp3⁺ T_{reg} cells to effector T cells. And if T_{reg} cells are transferred on their own, Foxp3-negative T cells that have formed from Foxp3⁺ T_{reg} cells produce IL-2, which inhibits the conversion of further Foxp3⁺ T_{reg} cells in a negative-feedback loop³. Similarly, in Rubtsov and colleagues' study², the sublethal X-ray radiation given might reduce T-cell levels substantially but not completely, allowing residual T cells to secrete IL-2 and thereby inhibit T_{reg} conversion.

Another possibility is that Foxp3⁺ T cells vary in functional stability and therefore in their susceptibility to conversion; that is, only a certain fraction of Foxp3⁺ T_{reg} cells may be 'plastic'. In support of this notion, the regulatory regions of the *Foxp3* gene are more extensively demethylated in thymus-derived T_{reg} cells than in those induced by TGF- β ¹⁰, offering an explanation for the functional instability of the latter population and its higher susceptibility to conversion. Even thymus-derived Foxp3⁺ T_{reg} cells could vary in the methylation status of their *Foxp3* gene, in their suppressive activity and in their eventual fate. Indeed, there is evidence¹¹ for functional and phenotypic variability in human Foxp3⁺ T cells: Foxp3^{high}

cells are terminally differentiated to be highly suppressive, whereas some Foxp3^{low} cells are non-suppressive and can secrete effector cytokines.

Finally, alterations in the level of Foxp3 expression may affect the functional stability of T_{reg} cells and, possibly, their susceptibility to conversion. For example, mice that express a reporter dye along with Foxp3 in their T_{reg} cells show a slight reduction in Foxp3 expression and an increased susceptibility to autoimmune disease¹². This might explain the discrepancies in T_{reg}-cell conversion between the two studies^{2,6}, which used different genetic manipulation methods such that different levels of the reporter dye were expressed.

Thus, the plasticity of T cells to differentiate to and from Foxp3⁺ T_{reg} cells is elaborately controlled by factors internal and external to these cells. Collectively, these factors ensure a remarkably constant number of Foxp3⁺ T_{reg} cells in the immune system (about 10% of all T cells expressing the surface marker CD4), with a general increase only occurring at sites of inflammation. Further understanding of the molecular basis of functional stability and the

plasticity of T_{reg} and other T cells should facilitate safe and effective control of physiological and disease-associated immune responses. ■

Shimon Sakaguchi is in the Department of Experimental Pathology, Institute for Frontier Medical Sciences, Kyoto University, Kyoto 606-8507, Japan, and is also at the WPI Immunology Frontier Research Center, Osaka University, Japan.
e-mail: shimon@frontier.kyoto-u.ac.jp

1. Sakaguchi, S., Yamaguchi, T., Nomura, T. & Ono, M. *Cell* **133**, 775–787 (2008).
2. Rubtsov, Y. P. *et al.* *Science* **329**, 1667–1671 (2010).
3. Duarte, J. H. *et al.* *Eur. J. Immunol.* **39**, 948–955 (2009).
4. Tsuji, M. *et al.* *Science* **323**, 1488–1492 (2009).
5. Oldenhove, G. *et al.* *Immunity* **31**, 772–786 (2009).
6. Zhou, X. *et al.* *Nature Immunol.* **10**, 1000–1007 (2009).
7. Bettelli, E., Korn, T., Oukka, M. & Kuchroo, V. K. *Nature* **453**, 1051–1057 (2008).
8. Yang, X. O. *et al.* *Immunity* **29**, 44–56 (2008).
9. Wan, Y. Y. & Flavell, R. A. *Nature* **445**, 766–770 (2007).
10. Huehn, J., Polansky, J. K. & Hamann, A. *Nature Rev. Immunol.* **9**, 83–89 (2009).
11. Miyara, M. *et al.* *Immunity* **30**, 899–911 (2009).
12. Wing, K. *et al.* *Science* **322**, 271–275 (2008).

DIABETES

Podocytes lose their footing

Impaired insulin action, combined with its insufficient secretion, can cause diabetes. In a surprising extension of this notion, decreased insulin action in the kidney's podocyte cells may contribute to renal complications in diabetes.

CHRISTIAN RASK-MADSEN & GEORGE L. KING

The hormone insulin is best known for regulating glucose and fat metabolism in skeletal muscle, fat tissue and the liver. Impaired insulin action in these tissues is central to the development of type 2 diabetes, which can be debilitating or even fatal, owing to kidney failure or cardiovascular complications. During the past decade or so, responses to insulin have been identified in other cell types, where some of its effects are completely different from those regulating metabolism. Writing in *Cell Metabolism*, Welsh *et al.*¹ show that, in mice, loss of insulin action in podocytes — a specialized type of kidney cell — can cause changes in kidney structure and function that resemble the renal complications of human diabetes known as diabetic nephropathy.

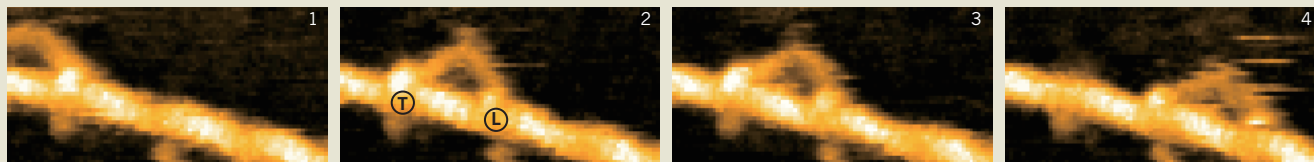
One of the kidney's main functions is to regulate water and salt balance in the body. To this end, the kidney performs ultrafiltration of blood plasma: every day, some 70 litres

of filtrate pass through the kidney's collecting ducts, while plasma macromolecules such as albumin are retained. Plasma filtration takes place through loops of capillaries in microscopic structures called renal glomeruli. These capillaries are covered by podocytes, which are epithelial cells that coat the entire capillary surface with the foot-like processes that give them their name. The filtration barrier consists of a single layer of endothelial cells — the cell type that forms the inner lining of all blood vessels — separated from the podocytes by a basal membrane.

One of the earliest abnormalities in the development of diabetic nephropathy is shortening of podocyte foot processes, which contributes to the breakdown of the glomerular filtration barrier, thereby allowing albumin and other macromolecules to escape into the urine. Albumin loss in the urine and other renal abnormalities, which in some cases progress to kidney failure, also frequently occur in patients with metabolic syndrome — a common condition, associated with obesity and

CELL BIOLOGY

Myosin in motion



Structural biologists are good at producing static snapshots of proteins, but seeing them in action is the ultimate goal. This is exactly what Kodera *et al.* have achieved in a remarkable study that appears elsewhere in this issue (N. Kodera, D. Yamamoto, R. Ishikawa and T. Ando *Nature* **468**, 72–76; 2010). They have been able to directly visualize myosin V, a cytoskeletal motor protein, as it ‘walks’ along actin filaments.

The authors have developed a high-speed atomic force microscopy (HS-AFM) method that enables them to generate rapid images of proteins at much higher resolution than light microscopy. Here they apply this technological advance to visualize the stepwise movement of myosin V. The protein consists of two heads connected by long necks (lever arms) to a globular tail domain through a coiled-coil helix. The tail binds to

various cargo molecules to transport them along the actin filaments, the energy required being generated by hydrolysis of ATP to ADP.

The HS-AFM images showing the leading (L) and trailing (T) heads of the tail-truncated motor can be seen in Figure 1a of the paper on page 73, part of which is reproduced here. In frame 1, myosin is advancing from the left; in frames 2 and 3, it has marched into full view, the two heads and necks now being visible; in frame 4, it has moved on a step. See go.nature.com/DHBOY9 for the full movie.

The authors’ analysis confirms the known behaviour of myosin motors, including the use of successive 36-nanometre steps along the actin filament and the ‘hand-over-hand’ movement of the motor heads. However, the study also provides convincing evidence for the hypothesized ‘lever-arm swing’. In this feature, tiny changes in

myosin’s head domain are amplified by the lever arm to produce large displacements at the far end of the neck that translate into movement of the whole protein along the actin filaments.

The new analysis also uncovers novel characteristics of the myosin V mechanism, such as a ‘stomping’ behaviour, in which either the L or T head becomes detached then rebinds to actin. The stomp is observed more frequently for the L than for the T head, but the T-head stomp often leads to a forward movement on the actin filament.

The insight into the behaviour of myosin V that Kodera *et al.* reveal will have a major impact on understanding the mechanisms involved in biological molecular motors. More generally, the technological advance provided by HS-AFM looks set to take a prominent place in the field of biomolecular imaging. **Deepa Nath**

impaired insulin action. Drugs that decrease blood glucose, lower blood pressure or inhibit the actions of the hormone angiotensin can delay, but not eliminate, the onset of diabetic nephropathy.

To relay its signal from the bloodstream to a cell’s interior, insulin binds to receptors on the cell surface. Welsh *et al.*¹ used mice in which the gene encoding the insulin receptor was deleted specifically from podocytes. At birth, kidney appearance in these animals was normal. At five weeks of age, however, they began to show excretion of albumin in the urine, shortening of the podocyte foot processes, increased deposition of components of the basal membrane, and a higher frequency of programmed podocyte death through apoptosis. Some animals even developed shrunken kidneys with prevalent scar tissue similar in appearance to the kidneys of humans with late-stage diabetic nephropathy.

These findings, however, are not just notable for their striking similarity to the pathology of diabetic nephropathy in humans. Welsh and colleagues’ mice also showed mild worsening of kidney function. This observation is intriguing because, in the most commonly studied rodent model of diabetes, destruction of insulin-producing cells with the drug streptozotocin causes no significant change in kidney function, despite resulting in microscopic

kidney abnormalities and albumin excretion in the urine².

Welsh *et al.* also report that insulin can reorganize the actin cytoskeleton in podocytes maintained in culture. This phenomenon resembles the actin remodelling seen when insulin causes translocation of glucose-transport proteins to the cell surface in fat or skeletal muscle cells³. Exactly how regulation of the cytoskeleton affects both podocyte foot processes and the filtration barrier requires more detailed investigation.

Remodelling of the actin cytoskeleton also does not explain the increased apoptosis of podocytes lacking insulin receptors. In this regard, the observation⁴ that insulin enhances the expression of the protein vascular endothelial growth factor (VEGF) — a crucial survival factor as well as a regulator of blood-vessel formation — might be of relevance. Podocytes are the main source of VEGF in the kidney: mice lacking VEGF specifically in podocytes show partial loss of all major cell types in the glomerulus, including podocytes^{5,6}. Moreover, VEGF expression is reduced in tissues such as heart muscle in animals with diabetes⁷. Insulin can also prevent apoptosis by other mechanisms^{8,9}, including inactivation of the transcription factor FoxO and inhibition of caspase-9, a signalling molecule that promotes apoptosis.

Whether insulin signalling to other cell types in the glomerulus is essential for maintenance of the filtration barrier is not known. Dysfunction of endothelial cells in the systemic circulation is associated with the initiation and progress of diabetic nephropathy, and endothelial cells respond to insulin by changing the production of factors that regulate blood-vessel tone and by decreasing oxidant production and increasing levels of antioxidant enzymes¹⁰.

It also remains to be seen whether podocytes, or other renal cells, are insulin resistant in diabetes and metabolic syndrome in other animal models and in humans. It could be that some cell types or insulin-signalling pathways are more susceptible to insulin resistance than others. For example, insulin increases sodium transport in the kidney’s tubular cells, but this aspect of its function is not affected in diabetes¹¹. To understand whether insulin resistance in other renal cells contributes to diabetic nephropathy, researchers must study normal insulin action in kidney cells; whether these cells develop insulin resistance in metabolic syndrome and/or in diabetes; and what causes impaired insulin signalling.

With its focus on insulin resistance in glomerular cells, Welsh and co-workers’ paper¹ helps to establish that diabetic nephropathy — the leading cause of chronic kidney

disease in developed countries — may result from the impaired actions of survival factors such as insulin, in addition to the deleterious effects of increased concentrations of glucose, angiotensin and other factors. This may provide the rationale for evaluating existing insulin-sensitizing drugs, as well as those under development, for their ability to improve insulin action in the kidney. It may also prove fruitful to develop insulin analogues that preferentially activate insulin-stimulated mechanisms in podocytes and other kidney cells. It is to be hoped that these strategies will provide

further ways to decrease the risk of diabetic nephropathy. ■

Christian Rask-Madsen and George L. King are in the Joslin Diabetes Center, Harvard Medical School, Boston, Massachusetts 02215, USA.

e-mail: george.king@joslin.harvard.edu

1. Welsh, G. I. *et al. Cell Metab.* **12**, 329–340 (2010).
2. Breyer, M. D. *et al. J. Am. Soc. Nephrol.* **16**, 27–45 (2005).
3. Kanzaki, M. *Endocr. J.* **53**, 267–293 (2006).

4. Zelzer, E. *et al. EMBO J.* **17**, 5085–5094 (1998).
5. Eremina, V. *et al. J. Clin. Invest.* **111**, 707–716 (2003).
6. Eremina, V. *et al. J. Am. Soc. Nephrol.* **17**, 724–735 (2006).
7. Chou, E. *et al. Circulation* **105**, 373–379 (2002).
8. Fu, Z. & Tindall, D. J. *Oncogene* **27**, 2312–2319 (2008).
9. Hermann, C., Assmus, B., Urbich, C., Zeiher, A. M. & Dimmeler, S. *Arterioscler. Thromb. Vasc. Biol.* **20**, 402–409 (2000).
10. Rask-Madsen, C. & King, G. L. *Nature Clin. Pract. Endocrinol. Metab.* **3**, 46–56 (2007).
11. Tiwari, S., Riaz, S. & Ecelbarger, C. A. *Am. J. Physiol. Renal. Physiol.* **293**, F974–F984 (2007).

QUANTUM COMPUTING

Quantum RAM

Hybrid quantum systems have been suggested as a potential route to building a quantum computer. The latest research shows that they offer a robust solution to developing a form of random access memory for such a machine.

MILES BLENCOWE

Most of us have shared the frustration of a desktop computer grinding almost to a halt when running a data-intensive application — opening a high-resolution digital photograph, for example — or running one application too many at the same time. Some have also experienced the (usually short-lived) improvement in speed that comes from installing expensive additional memory called random access memory (RAM). Unlike that in the hard drive, data stored in RAM can be retrieved just as quickly in any order, making it well suited for its role as a temporary storage medium for the computer's central processing unit while executing a program. A quantum computer will also require a form of RAM for its proper function. Writing in *Physical Review B*¹ and in *Physical Review Letters*^{2–4}, four research groups report significant progress in demonstrating a proof-of-principle 'quantum RAM'.

As with a conventional computer, a quantum computer encodes the binary digits 0 and 1 — that is, a 'bit' of information — in the state of a physical system. But in contrast to its classical counterpart, it does not do so using the state of a classical system, such as the absence or presence of an electrical charge in a capacitor. The 0 and 1 will correspond instead to the two states of a two-level quantum system — for example, the spin 'down' and spin 'up' states of unpaired electrons of atomic or molecular defects in a crystal lattice, or the clockwise and anticlockwise electrical currents in a tiny superconducting ring. And because of quantum mechanics, both of these example quantum systems can be not only in either the 0 or 1 state, but also in a superposition state — a

simultaneous combination of 0 and 1. As such, they act as a 'quantum bit', or qubit, to encode quantum information.

In their papers^{1–4}, the four teams describe work towards realizing a hybrid quantum-computer architecture that combines qubits

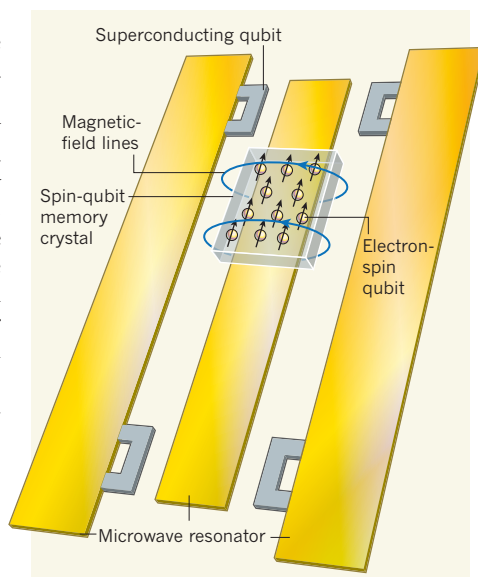


Figure 1 | Scheme for a hybrid quantum computer. The hybrid quantum-computer architecture envisaged in four papers^{1–4} combines superconducting qubits and spin qubits. The superconducting-qubit states can be transferred to and retrieved from a spin-qubit memory crystal by microwave-resonator photons. The microwave resonator strongly couples to the crystal's ensemble of electron-spin memory qubits through the tightly confined oscillating magnetic field of the resonator photons. The resonator photons can also mediate couplings between the superconducting qubits to realize quantum logic gates⁶.

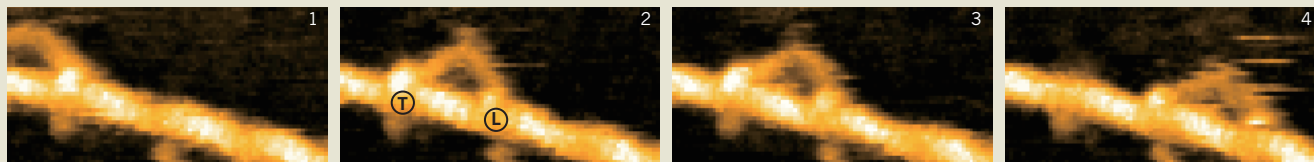
of both types described above: superconducting qubits and spin qubits (Fig. 1). The superconducting qubits, which are typically a few hundred micrometres in size, are well suited to performing fast quantum logic-gate operations^{5,6}. In addition, they are relatively straightforward to fabricate from materials such as aluminium using electron-beam lithography. The spin qubits, which are formed from a millimetre-to-centimetre-sized crystal containing upwards of 10^{12} electron-spin impurities, act as memory elements to store and retrieve data. The spin qubits considered by the authors include chromium(III) ions (Cr^{3+}) in aluminium oxide², nitrogen-vacancy defects in diamond³, and rather exotic molecules consisting of single nitrogen atoms in fullerene (C_{60}) cages⁴ (the fullerene cage prevents the normally chemically unstable nitrogen atoms from reacting). Such spin-qubit memories have the advantage of relatively long (millisecond) lifetimes, after which the stored information will typically have decayed away owing to unavoidable interactions with the uncontrolled environment of the spins. This lifetime is more than a thousand times longer than the microsecond or shorter lifetimes of the superconducting qubits.

In such hybrid schemes, the data buses that transport information between computer components can be fashioned from long, thin metallic strips of aluminium or some other suitable superconductor such as niobium. The strips can be carefully engineered to form microwave resonators that have resonant frequencies in the few-to-several gigahertz range, enabling the superconducting (logic) and spin (memory) qubits to emit and absorb resonant-frequency microwave photons and hence exchange information between one another.

The new studies^{1–4} do not demonstrate such a hybrid quantum-computer architecture in its entirety. What the authors show is a strong coupling between the microwave resonator and the ensemble of electron-spin memory qubits in the crystal. To realize a quantum RAM, the coupling must be sufficiently strong for a microwave-resonator photon to be stored in the spin-qubit memory and retrieved on a timescale that is short

CELL BIOLOGY

Myosin in motion



Structural biologists are good at producing static snapshots of proteins, but seeing them in action is the ultimate goal. This is exactly what Kodera *et al.* have achieved in a remarkable study that appears elsewhere in this issue (N. Kodera, D. Yamamoto, R. Ishikawa and T. Ando *Nature* **468**, 72–76; 2010). They have been able to directly visualize myosin V, a cytoskeletal motor protein, as it ‘walks’ along actin filaments.

The authors have developed a high-speed atomic force microscopy (HS-AFM) method that enables them to generate rapid images of proteins at much higher resolution than light microscopy. Here they apply this technological advance to visualize the stepwise movement of myosin V. The protein consists of two heads connected by long necks (lever arms) to a globular tail domain through a coiled-coil helix. The tail binds to

various cargo molecules to transport them along the actin filaments, the energy required being generated by hydrolysis of ATP to ADP.

The HS-AFM images showing the leading (L) and trailing (T) heads of the tail-truncated motor can be seen in Figure 1a of the paper on page 73, part of which is reproduced here. In frame 1, myosin is advancing from the left; in frames 2 and 3, it has marched into full view, the two heads and necks now being visible; in frame 4, it has moved on a step. See go.nature.com/DHBOY9 for the full movie.

The authors’ analysis confirms the known behaviour of myosin motors, including the use of successive 36-nanometre steps along the actin filament and the ‘hand-over-hand’ movement of the motor heads. However, the study also provides convincing evidence for the hypothesized ‘lever-arm swing’. In this feature, tiny changes in

myosin’s head domain are amplified by the lever arm to produce large displacements at the far end of the neck that translate into movement of the whole protein along the actin filaments.

The new analysis also uncovers novel characteristics of the myosin V mechanism, such as a ‘stomping’ behaviour, in which either the L or T head becomes detached then rebinds to actin. The stomp is observed more frequently for the L than for the T head, but the T-head stomp often leads to a forward movement on the actin filament.

The insight into the behaviour of myosin V that Kodera *et al.* reveal will have a major impact on understanding the mechanisms involved in biological molecular motors. More generally, the technological advance provided by HS-AFM looks set to take a prominent place in the field of biomolecular imaging. **Deepa Nath**

impaired insulin action. Drugs that decrease blood glucose, lower blood pressure or inhibit the actions of the hormone angiotensin can delay, but not eliminate, the onset of diabetic nephropathy.

To relay its signal from the bloodstream to a cell’s interior, insulin binds to receptors on the cell surface. Welsh *et al.*¹ used mice in which the gene encoding the insulin receptor was deleted specifically from podocytes. At birth, kidney appearance in these animals was normal. At five weeks of age, however, they began to show excretion of albumin in the urine, shortening of the podocyte foot processes, increased deposition of components of the basal membrane, and a higher frequency of programmed podocyte death through apoptosis. Some animals even developed shrunken kidneys with prevalent scar tissue similar in appearance to the kidneys of humans with late-stage diabetic nephropathy.

These findings, however, are not just notable for their striking similarity to the pathology of diabetic nephropathy in humans. Welsh and colleagues’ mice also showed mild worsening of kidney function. This observation is intriguing because, in the most commonly studied rodent model of diabetes, destruction of insulin-producing cells with the drug streptozotocin causes no significant change in kidney function, despite resulting in microscopic

kidney abnormalities and albumin excretion in the urine².

Welsh *et al.* also report that insulin can reorganize the actin cytoskeleton in podocytes maintained in culture. This phenomenon resembles the actin remodelling seen when insulin causes translocation of glucose-transport proteins to the cell surface in fat or skeletal muscle cells³. Exactly how regulation of the cytoskeleton affects both podocyte foot processes and the filtration barrier requires more detailed investigation.

Remodelling of the actin cytoskeleton also does not explain the increased apoptosis of podocytes lacking insulin receptors. In this regard, the observation⁴ that insulin enhances the expression of the protein vascular endothelial growth factor (VEGF) — a crucial survival factor as well as a regulator of blood-vessel formation — might be of relevance. Podocytes are the main source of VEGF in the kidney: mice lacking VEGF specifically in podocytes show partial loss of all major cell types in the glomerulus, including podocytes^{5,6}. Moreover, VEGF expression is reduced in tissues such as heart muscle in animals with diabetes⁷. Insulin can also prevent apoptosis by other mechanisms^{8,9}, including inactivation of the transcription factor FoxO and inhibition of caspase-9, a signalling molecule that promotes apoptosis.

Whether insulin signalling to other cell types in the glomerulus is essential for maintenance of the filtration barrier is not known. Dysfunction of endothelial cells in the systemic circulation is associated with the initiation and progress of diabetic nephropathy, and endothelial cells respond to insulin by changing the production of factors that regulate blood-vessel tone and by decreasing oxidant production and increasing levels of antioxidant enzymes¹⁰.

It also remains to be seen whether podocytes, or other renal cells, are insulin resistant in diabetes and metabolic syndrome in other animal models and in humans. It could be that some cell types or insulin-signalling pathways are more susceptible to insulin resistance than others. For example, insulin increases sodium transport in the kidney’s tubular cells, but this aspect of its function is not affected in diabetes¹¹. To understand whether insulin resistance in other renal cells contributes to diabetic nephropathy, researchers must study normal insulin action in kidney cells; whether these cells develop insulin resistance in metabolic syndrome and/or in diabetes; and what causes impaired insulin signalling.

With its focus on insulin resistance in glomerular cells, Welsh and co-workers’ paper¹ helps to establish that diabetic nephropathy — the leading cause of chronic kidney

compared with the photon lifetime in the resonator and the spin-qubit memory lifetime.

In their experiments, the authors^{1–4} achieve strong coupling in two ways. First, the large cross-sectional aspect ratio (the ratio of width to height) of the wide but thin microwave resonator strip means that the photon's oscillating magnetic-field component, which interacts with the electron spins, is largely confined to a small volume above and below the strip, comparable to the thickness of the spin-qubit memory crystal. Second, the microwave photon's centimetre-scale wavelength along the strip length is comparable to the lateral dimensions of the crystal. Thus, with the crystal positioned on top of the strip, a significant portion of the photon's magnetic field will fill the crystal volume; a single photon therefore interacts simultaneously with a large number of the spins, with the result that the photon is stored non-locally throughout the crystal in the form of a many-spin superposition state. A strong coupling then depends simply on ensuring a high crystal spin density. But how is this coupling strength actually measured?

Consider two pendulums suspended from a ceiling with strings of identical length and bobs of identical mass. Suppose one of the pendulums represents the microwave resonator and the other the non-local spin 'mode'. If the two independent pendulums are set into equal, but small-amplitude motion, then they will oscillate at the same frequency — that is, in resonance. Now attach a spring between the two masses, coupling their motion. The spring represents the coupling between the microwave resonator and the spin mode. If we then initially displace one of the pendulums, it will gradually transfer all of its energy to the other pendulum and vice versa, resulting in a 'beat' frequency — this is similar to what is heard when two musical notes are played slightly out of tune with respect to each other. The beat frequency is directly related to the spring stiffness, and hence the coupling strength, between the pendulums. Analogously, the authors^{1–4} measured the beat frequency of the microwave resonator when in resonance with the spin mode, and hence were able to extract their coupling strength.

Only one of the four experiments actually demonstrated⁴ the RAM principle of multiple-state storage followed by arbitrary-order retrieval of the states. However, only classical microwave states involving a large, indeterminate number of photons were stored in and retrieved from the spin memory. To demonstrate the same at the single-photon level — as required for a quantum RAM — will require about an order-of-magnitude improvement in the resonator photon and spin-qubit memory lifetimes, together with the introduction of the superconducting qubits. Nevertheless, the current results represent a major step towards realizing a quantum RAM. ■

Miles Blencowe is in the Department of Physics and Astronomy, 6127 Wilder Laboratory, Dartmouth College, Hanover, New Hampshire 03755, USA.
e-mail: miles.p.blencowe@dartmouth.edu

1. Chiorescu, I., Groll, N., Bertaina, S., Mori, T. &

Miyashita, S. *Phys. Rev. B* **82**, 024413 (2010).
2. Schuster, D. I. et al. *Phys. Rev. Lett.* **105**, 140501 (2010).
3. Kubo, Y. et al. *Phys. Rev. Lett.* **105**, 140502 (2010).
4. Wu, H. et al. *Phys. Rev. Lett.* **105**, 140503 (2010).
5. Neeley, M. et al. *Nature* **467**, 570–573 (2010).
6. DiCarlo, L. et al. *Nature* **467**, 574–578 (2010).

MICROBIOLOGY

Slicer for DNA

Many bacteria and archaea protect themselves from viruses and other invasive genomes through a genetic interference pathway. The small RNAs that guide this defence specify the direct cleavage of foreign DNA. SEE ARTICLE P.67

**ERIK J. SONTHEIMER
& LUCIANO A. MARRAFFINI**

All cells invest heavily in maintaining their genetic identities against invasion by foreign nucleic acids such as viral genomes. This problem is particularly acute for bacteria and archaea because of the pervasive presence of bacteriophages, plasmids and other mobile genetic elements in the biosphere. Genetic entities called clustered regularly interspaced short palindromic repeats (CRISPRs) have been revealed as adaptive, genomically

encoded immune systems¹ that protect bacteria and archaea from phages², plasmids³ and probably other forms of foreign DNA.

This system is sequence-directed, like the well-known RNA interference (RNAi) pathway⁴ that operates in organisms such as plants and animals. If a portion of the CRISPR locus matches a sequence from the invasive genome, the invader is thwarted. This mechanism requires the action of CRISPR RNAs (crRNAs) that render the pathway addressable⁵, presumably by delivering an interference machinery to target nucleic acids that are recognized by the

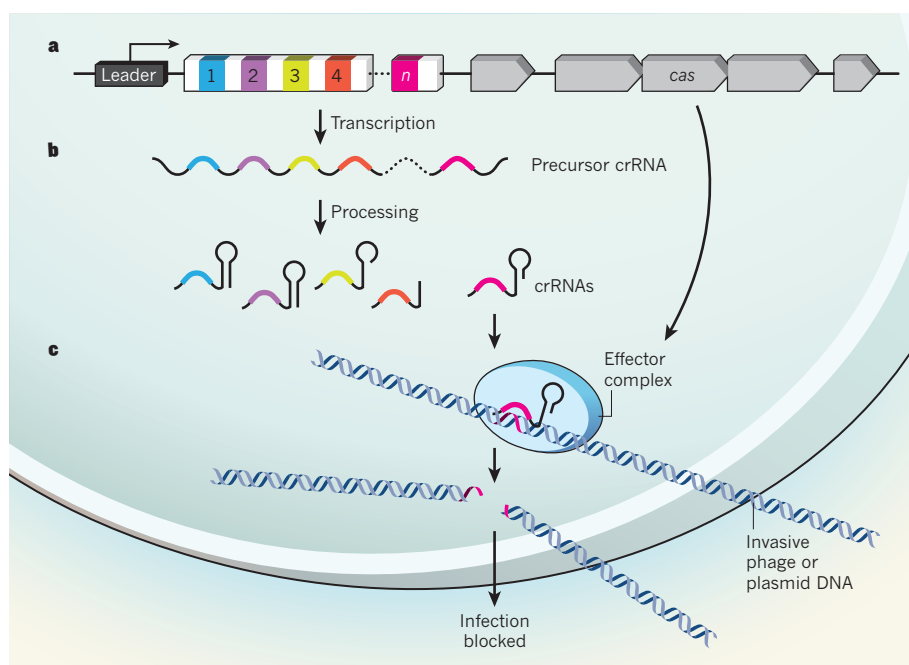


Figure 1 | RNA-guided DNA cleavage by the core CRISPR machinery. **a**, CRISPR loci include variable numbers of repeat and spacer sequences, with the latter derived from invasive DNAs previously encountered by the cell or its ancestors. The repeat/spacer region is flanked by a 'leader' sequence and by *cas* genes¹. **b**, CRISPR transcription and processing of precursor crRNA yields mature crRNAs. **c**, These presumably associate with an effector complex that includes protein products of the *cas* genes. Garneau et al.⁶ report that the CRISPR pathway in *Streptococcus thermophilus* induces double-strand breaks at cognate protospacers embedded within invasive DNAs, implying that the core effector complex achieves genetic interference, and so infection blockage, by crRNA-guided DNA cleavage.

complementarity of Watson–Crick base pairing. On page 67 of this issue, Garneau *et al.*⁶ now show that the core CRISPR machinery protects cells in the most direct way conceivable: by crRNA-specified cleavage of the foreign DNA (Fig. 1).

The basic features of any CRISPR system include the CRISPR locus itself, which consists of short repeats that are separated by non-repetitive sequences called spacers. These spacers match 'proto-spacer' sequences from phage genomes, plasmids and other previously encountered foreign DNAs, providing a genomically recorded memory of past invasions. This memory is expressed in the form of crRNAs. CRISPR loci are also flanked by a set of CRISPR-associated (*cas*) genes that encode protein components of the interference machinery.

When cells are challenged by a phage with no match to any CRISPR spacer, they are unprotected and (if no other anti-phage defence pathways are operative) die. However, a small fraction of infected cells can hijack pieces of the phage genome for the purpose of 'immunization': one or a few new phage-derived spacers are inserted at one end of the CRISPR locus, conferring immunity to that phage and allowing the adapted cell and its progeny to flourish even in the face of repeated attacks.

After the CRISPR pathway was uncovered as an adaptive, RNA-directed immune system², an immediate issue was the identity of the molecular target. Does CRISPR interference (like RNAi) target RNAs transcribed from invasive genomes, or does it target invasive DNA itself? DNA was quickly recognized as a prime suspect in the bacterium *Escherichia coli*⁵ because crRNAs were found to be single-stranded entities without an antisense bias, making it difficult (but not impossible)⁷ to envisage RNA as the target. More definitive evidence for DNA targeting was reported shortly thereafter in *Staphylococcus epidermidis*³. In spite of the genetic evidence for DNA recognition, however, the precise molecular mechanism of interference was still not proven. Formal possibilities included DNA cleavage (the leading contender), DNA sequestration, and inhibition of any number of essential DNA-based processes, such as transcription or replication.

Enter Garneau and colleagues⁶. In their efforts to establish that plasmids can induce CRISPR adaptation and interference in *Streptococcus thermophilus*, they made a surprising observation: one adapted strain with only partial plasmid resistance accumulated a linear plasmid form that was easily lost under non-selective conditions. Only the circular form was found in non-adapted strains or in a derivative of the adapted strain with a key *cas* gene disrupted. Sequencing of the linear plasmid revealed a break at a specific site near one end of the matching protospacer, suggesting a crRNA-based measuring mechanism.

This phenomenon applies to phage

interference as well, because Southern blotting revealed the presence of phage genome fragments in multiple infected strains carrying matching CRISPR spacers. Crucially, when strains carrying distinct CRISPR spacers were tested, the cleavage site moved in concert with the site of the cognate protospacer in the phage genome. During generation of sequencing templates for cleavage-site mapping, the protospacer ends that were cleaved *in vivo* could ligate to an opposite blunt end of the DNA fragment generated by a restriction enzyme, indicating that CRISPR cleavage generates blunt ends as well, in agreement with the sequencing data.

Multiple categories of CRISPR/Cas loci exist with overlapping but distinct sets of Cas proteins¹, raising the likelihood of mechanistic differences between the subtypes. It is likely that DNA cleavage is a general mechanism that does not apply only to *S. thermophilus*, but such generality remains to be established. Intriguingly, proteins encoded by 'RAMP module' genes, which accompany core CRISPR/Cas loci in a subset of mostly archaeal species, have been shown to catalyse crRNA-guided RNA cleavage *in vitro*⁷. Thus, as with RNAi, multiple branches of the CRISPR interference pathway probably coexist. It will be a fascinating task to disentangle the *in vivo* relationships between the RNA-targeting RAMP module and the DNA-targeting core Cas machinery.

With crRNA-directed double-stranded DNA cleavage now established, the race is on to identify and characterize the DNA cleavage effector enzymes in *S. thermophilus* as well as representatives of other CRISPR/Cas subtypes.

STRUCTURAL BIOLOGY

On stress and pressure

The protein angiotensinogen must undergo conformational changes to be cleaved into a precursor of the hormone angiotensin, which increases blood pressure. Oxidative stress seems to mediate this structural alteration. SEE LETTER P.108

CURT D. SIGMUND

Numerous signals emanating from the kidney, blood vessels, heart and brain regulate blood pressure, with an imbalance between them causing altered blood pressure. One such prominent signalling pathway is the renin–angiotensin system. Renin is a protease enzyme that catalyses the rate-limiting step in this pathway — cleavage of ten amino-acid residues at the amino terminus of its only known substrate, angiotensinogen. The resulting peptide is further processed into a family of angiotensin peptide hormones, the best characterized of which, angiotensin II, increases blood pressure in various ways.

A strong suspect (Cas3) with a likely double-strand-specific HD nuclease domain⁸ has already been implicated in *E. coli*⁵. Interestingly, the CRISPR/Cas loci in *S. thermophilus*⁶ and *S. epidermidis*³ — where the evidence for DNA targeting is strongest — do not include obvious *cas3* genes, indicating that multiple, subtype-specific effectors remain to be identified.

Finally, Garneau and colleagues' demonstration⁶ of crRNA-directed DNA cleavage could have practical as well as biological import. As we have noted previously³, the potential to specify RNA-guided, sequence-specific DNA cleavage via simple Watson–Crick pairing rules, rather than through complex protein–DNA interactions, opens up many enticing practical possibilities. ■

Erik J. Sontheimer is in the Department of Molecular Biosciences, Northwestern University, Evanston, Illinois 60208, USA.

Luciano A. Marraffini is in the Laboratory of Bacteriology, Rockefeller University, New York, New York 10065, USA.

e-mails: erik@northwestern.edu; marraffini@mail.rockefeller.edu

1. Karginov, F. V. & Hannon, G. J. *Mol. Cell* **37**, 7–19 (2010).
2. Barrangou, R. *et al.* *Science* **315**, 1709–1712 (2007).
3. Marraffini, L. A. & Sontheimer, E. J. *Science* **322**, 1843–1845 (2008).
4. Liu, Q. & Paroo, Z. *Annu. Rev. Biochem.* **79**, 295–319 (2010).
5. Brouns, S. J. J. *et al.* *Science* **321**, 960–964 (2008).
6. Garneau, J. E. *et al.* *Nature* **468**, 67–71 (2010).
7. Hale, C. R. *et al.* *Cell* **139**, 945–956 (2009).
8. Han, D. & Krauss, G. *FEBS Lett.* **583**, 771–776 (2009).

On page 108 of this issue, Zhou *et al.*¹ provide structural evidence for the way in which redox-based conformational changes in angiotensinogen expose the site cleaved by renin.

The renin–angiotensin system is evolutionarily 'thrifty'²: the common outcome of angiotensin activity is sodium conservation. In an ancient environment in which sodium was scarce, angiotensin activity may have been beneficial. However, the system can have detrimental effects in the modern, sodium-replete environment. Indeed, the sodium-preserving actions of angiotensin are problematic for the large population of salt-sensitive people with high blood pressure (hypertension), and drug-mediated blockade of the renin–angiotensin

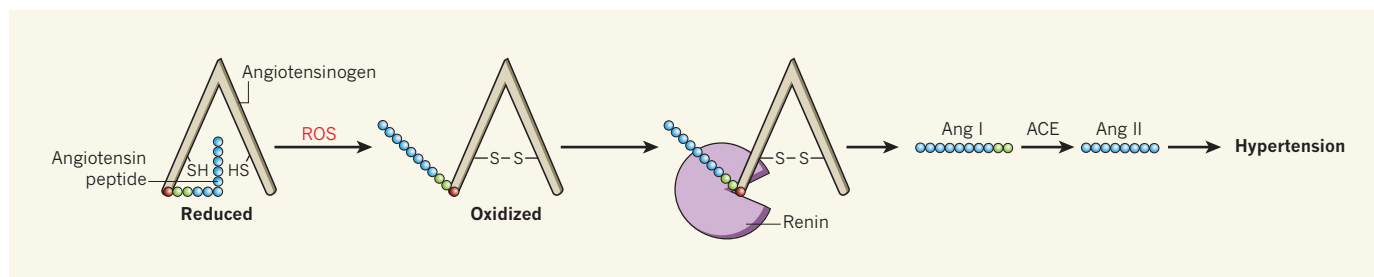


Figure 1 | The renin–angiotensin system. Zhou *et al.*¹ show that in the reduced form of angiotensinogen, the angiotensin peptide and the peptide's cleavage site (red/green) are inaccessibly buried. Reactive oxygen species (ROS) cause angiotensinogen oxidation and disulphide-bridge formation between the thiol (SH) groups of its C18 and C138 residues. This results in a

conformational change that makes the angiotensin peptide and its cleavage site accessible to renin. The initial product formed by the renin–angiotensinogen cleavage reaction is the ten-amino-acid peptide angiotensin I (Ang I), which can be cleaved further by angiotensin converting enzyme (ACE) to angiotensin II (Ang II), a hormone commonly associated with hypertension.

system is of clinical importance because it can be an effective treatment for many people with hypertension³.

Although angiotensin can activate diverse cellular processes on binding to its cellular receptors, a common thread is the generation of reactive oxygen species⁴. For instance, activation of angiotensin receptors can promote the generation of superoxide and induces oxidative stress in the main cardiovascular target organs such as the kidney, blood vessels and brain. (Oxidative stress results from an imbalance between the production of reactive oxygen species and the system's ability to readily detoxify reactive intermediates.) Superoxide is a highly reactive free radical that can be converted to lipid-soluble reactive products, including hydrogen peroxide and peroxynitrite. The link between the renin–angiotensin system and superoxide has been established both pharmacologically and genetically. In fact, treatment with antioxidants can lower blood pressure in rat and mouse models of angiotensin-induced hypertension. Such findings further support the pro-hypertensive effects of reactive oxygen species as a downstream effector of the renin–angiotensin system.

One consequence of oxidative stress is chemical modification of proteins at cysteine amino-acid residues. Interaction between hydrogen peroxide and the thiol group on cysteines results in the formation of a sulphenic-acid intermediate that can react with nearby cysteines to form disulphide bonds, altering a protein's structure and activity. But what is the link between oxidative alterations to proteins and hypertension? Zhou *et al.*¹ provide a glimpse of the answer.

The authors present the crystal structure of angiotensinogen at a resolution of 2.1 Å and show that the cleavage site for angiotensin generation is inaccessibly buried inside the protein's amino-terminal tail (Fig. 1). To identify the conformational changes required to make this site accessible to renin, they also solved the crystal structure of the initiating complex. This consisted of genetically engineered angiotensinogen and a catalytically inactive mutant of renin. The authors' analyses reveal that angiotensinogen's amino-terminal

tail — including the angiotensin cleavage site — undergoes a conformational shift that makes it accessible to renin. The linchpin of this structural change seems to be the formation of a disulphide bridge between a pair of evolutionarily conserved cysteine residues (C18 and C138) in angiotensinogen.

The oxidized (disulphide-bridged) state of angiotensinogen correlated with the protein's increased affinity for renin, and so increased the production of angiotensin¹. Intriguingly, these increases were particularly evident when renin was in complex with the prorenin receptor, a membrane-bound renin receptor of obscure physiological significance that has been reported⁵ to increase the catalytic activity of both renin and its inactive precursor prorenin.

Notably, Zhou *et al.* find that the C18–C138 disulphide bridge is redox sensitive: the anti-oxidant molecule glutathione breaks the sulphur–sulphur link to regenerate two thiol groups. This may be physiologically significant because endogenous nitric oxide, which is a potent dilator of blood vessels, thus reducing blood pressure, can alter cysteine thiols by means of S-nitrosylation (addition of a nitrogen monoxide group). Zhou *et al.* find that a nitric-oxide donor reacts with and blocks the thiols on C18 and C138, thereby possibly preventing further oxidation of the cysteine residues. This finding suggests that intracellular chemicals such as reactive oxygen species and nitric oxide can tightly regulate the redox state of angiotensinogen — and so its processing — in tissues. The implication is that the rate of angiotensin generation may be as much dictated by the redox states of the tissue and blood as by the absolute levels of renin and angiotensinogen.

Under normal conditions, there is roughly a 40:60 ratio of reduced-to-oxidized angiotensinogen in the circulation. Are the levels of oxidized angiotensinogen increased in hypertension? Yes, according to Zhou and colleagues. They demonstrate that the proportion of oxidized angiotensinogen was significantly higher in the blood plasma of 12 patients with pre-eclampsia — a syndrome of hypertension and increased levels of serum proteins in the

urine during pregnancy. This is interesting in light of previous findings that a variant form of the gene encoding angiotensinogen is associated with hypertension⁶ and pre-eclampsia⁷.

We must now consider whether Zhou and co-workers' structural findings are applicable to common forms of hypertension. The investigators' observations suggest that angiotensin-induced production of reactive oxygen species results in a feed-forward mechanism favouring an ever-increasing activity of the renin–angiotensin system and its pro-hypertensive effects. Can this cycle be amplified by genetic variants that either promote additional angiotensinogen synthesis or favour its structurally active form?

The most commonly studied variant within the protein-coding regions of the angiotensinogen gene is located away from the functional moieties that Zhou *et al.*¹ identify. However, this variant is linked to another one found in the promoter region of the angiotensinogen gene that enhances angiotensinogen production^{8,9}. My laboratory has just shown¹⁰ in mice that the presence of both variants affects blood pressure only in animals on a high-salt diet — a condition that promotes oxidative stress. Therefore, given what is now known about the oxidation-associated changes in angiotensinogen, it would be intriguing to assess whether variants in the angiotensinogen gene affect the protein's redox-regulated activity under conditions that promote oxidative stress. ■

Curt D. Sigmund is in the Department of Pharmacology, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, Iowa 52242, USA.
e-mail: curt-sigmund@uiowa.edu

1. Zhou, A. *et al.* *Nature* **468**, 108–111 (2010).
2. Sharma, A. M. *J. Mol. Med.* **76**, 568–571 (1998).
3. Chobanian, A. V. *et al.* *Hypertension* **42**, 1206–1252 (2003).
4. Rajagopalan, S. *et al.* *J. Clin. Invest.* **97**, 1916–1923 (1996).
5. Nguyen, G. *et al.* *J. Clin. Invest.* **109**, 1417–1427 (2002).
6. Jeunemaitre, X. *et al.* *Cell* **71**, 169–180 (1992).
7. Ward, K. *et al.* *Nature Genet.* **4**, 59–61 (1993).
8. Morgan, T., Craven, C., Nelson, L., Lalouel, J. M. & Ward, K. *J. Clin. Invest.* **100**, 1406–1415 (1997).
9. Inoue, I. *et al.* *J. Clin. Invest.* **99**, 1786–1797 (1997).
10. Grobe, J. L. *et al.* *Hypertension* **56**, 981–987 (2010).

Early star-forming galaxies and the reionization of the Universe

Brant E. Robertson¹, Richard S. Ellis¹, James S. Dunlop², Ross J. McLure² & Daniel P. Stark³

Star-forming galaxies trace cosmic history. Recent observational progress with the NASA Hubble Space Telescope has led to the discovery and study of the earliest known galaxies, which correspond to a period when the Universe was only ~800 million years old. Intense ultraviolet radiation from these early galaxies probably induced a major event in cosmic history: the reionization of intergalactic hydrogen.

The frontier in completing the physical story of cosmic history is to understand cosmic reionization—the transformation of neutral hydrogen, mostly located outside galaxies in the intergalactic medium (IGM), into an ionized state. Neutral hydrogen first formed 370,000 years after the Big Bang and released the radiation presently observed as the cosmic microwave background (CMB)¹. Initially devoid of sources of light, the Universe then entered a period termed the ‘Dark Ages’², which lasted until the first stars formed from overdense clouds of hydrogen gas that cooled and collapsed within early cosmic structures. Observations of distant quasars³ demonstrate that the IGM has been highly ionized since the Universe was ~1 billion years (Gyr) old, and the transition from a neutral medium is popularly interpreted as being caused by ionizing photons with energies greater than 13.6 eV (wavelength, $\lambda < 91.2$ nm), generated by primitive stars and galaxies⁴ (Fig. 1).

Astronomers wish to confirm the connection between early galaxies and reionization because detailed studies of this period of cosmic history will reveal the physical processes that originally shaped the galaxies of various luminosities and masses we see around us today. Alternative sources of reionizing photons include material collapsing into early black holes that power active galactic nuclei, and decaying elementary particles. Verifying that star-forming galaxies were responsible for cosmic reionization requires understanding how many energetic ultraviolet photons were produced by young stars at early times and what fraction of photons capable of ionizing hydrogen outside galaxies escaped without

being intercepted by clouds of dust and hydrogen within galaxies. Astronomers desire accurate measurements of the abundance of early galaxies and the distribution of their luminosities to quantify the number of sources producing energetic photons, as well as a determination of the mixture of stars, gas and dust in galaxies, to determine the likelihood that the ultraviolet radiation can escape to ionize the IGM^{5,6}. The Lyman- α (Ly α) emission line, which is detectable using spectrographs on large ground-based telescopes, is a valuable additional diagnostic given that it is easily erased by neutral gas outside galaxies^{7–12}. Its observed strength in distant galaxies is therefore a sensitive gauge of the latest time when reionization was completed.

In this primarily observational Review, we discuss substantial progress that now points towards a fundamental connection between early galaxies and reionization. Recent observations with the Hubble Space Telescope (HST) have provided the first detailed constraints on the abundance and properties of galaxies in the first 1 Gyr of cosmic history. With some uncertainties, these data indicate that sufficient ultraviolet radiation was produced to establish and maintain an ionized Universe by redshift $z \approx 7$, corresponding to ~800 million years (Myr) after the Big Bang. Further observations of these early systems using current facilities will produce a more robust census and clarify what fraction of the ionizing radiation escaped primitive galaxies. The rapid progress now being made will pave the way for ambitious observations of the earliest known galaxies using future facilities.

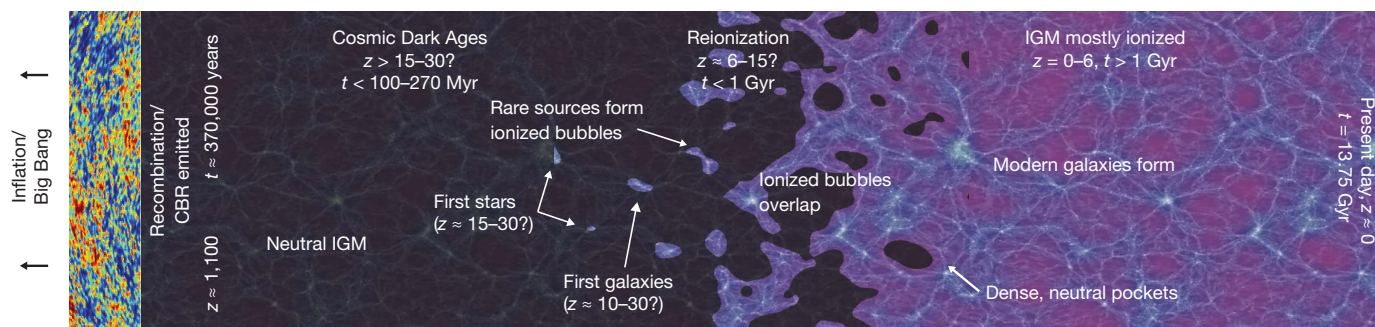


Figure 1 | Cosmic reionization. The transition from the neutral IGM left after the Universe recombined, at $z \approx 1,100$, to the fully ionized IGM observed today is termed cosmic reionization. After recombination, when the CMB radiation was released, hydrogen in the IGM remained neutral until the first stars and galaxies^{2,4} formed, at $z \approx 15–30$. These primordial systems released energetic ultraviolet photons capable of ionizing local bubbles of hydrogen gas. As the

abundance of these early galaxies increased, the bubbles increasingly overlapped and progressively larger volumes became ionized. This reionization process ended at $z \approx 6–8$, ~1 Gyr after the Big Bang. At lower redshifts, the IGM remains highly ionized by radiation provided by star-forming galaxies and the gas accretion onto supermassive black holes that powers quasars.

¹Department of Astronomy, California Institute of Technology, MC 249-17, Pasadena, California 91125, USA. ²Institute for Astronomy, University of Edinburgh, Edinburgh EH9 3HJ, UK. ³Institute of Astronomy, University of Cambridge, Cambridge CB3 0HA, UK.

Probes of the reionization epoch

Cosmological reionization involves a complex interplay between the strength, distribution and spectrum of photoionizing sources and the density and spatial structure of intergalactic gas. Theoretical calculations using both analytical methods^{13–17} and sophisticated numerical simulations^{18–28} to model these complexities have reached broad agreement on how reionization is expected to proceed (Fig. 1). Sources of ionizing photons, such as star-forming galaxies, are associated with dark matter halos forged through hierarchical structure formation²⁹. Initially, these sources are rare and only ionize sparse localized ‘bubbles’ in the dense IGM. As the Universe expands and the mean IGM density decreases, early sources grow in abundance, mass and luminosity with time and the ionized regions increase in number and extent. The ionized bubbles eventually overlap, allowing the mean free path of ionizing photons and the average volume fraction of ionized hydrogen, $Q_{\text{H II}}$, in the IGM to increase rapidly. The process is completed with the ionization of neutral pockets of hydrogen gas isolated from luminous sources. A technical description of the reionization process and the key requirements for star-forming galaxies to be the responsible agents are provided in Box 1.

The complete ionization of hydrogen in the IGM requires sustained sources of Lyman-continuum photons with wavelengths $\lambda < 91.2$ nm. If galaxies are responsible, the process of reionization should mirror their

time-dependent density. The number of Lyman-continuum photons produced by star-forming galaxies can be modelled and, assuming that some fraction can escape the galaxies, the average volume fraction of hydrogen ionized by these photons ($Q_{\text{H II}}$) can be calculated from the star formation rate density. The history of star formation is further constrained by its time integral, the mass density of long-lived stars. At high redshift, the presence of such stars is usefully probed by the NASA Spitzer Space Telescope^{30,31}. A final constraint is the optical depth for Thomson scattering by free electrons associated with reionization, as inferred by their ability to polarize CMB photons. Observational constraints on the electron optical depth³² suggest that an extended, low level of ionization ($\sim 10\%$ by volume) may be needed to high redshifts ($z > 20$) if reionization gradually ends at $z \approx 5.5$ – 8.5 . Reconciling the electron scattering optical depth with the number of available ionizing photons inferred from the currently observed star formation rate may therefore be difficult.

The above observations are discussed in more detail in Box 2. These indirect constraints are broadly consistent with a gradual reionization starting at $z \approx 20$ and ending at $z \approx 6$, as first indicated by measures of neutral hydrogen absorption in distant quasar and gamma-ray burst spectra that can probe the end of reionization^{33–36}. Ultimately, we may chart the distribution of ionized bubbles forming within intergalactic neutral hydrogen directly using radio interferometers sensitive to the hyperfine transition at 21 cm, but this topic is not covered explicitly here (for a review, see ref. 37). Such technically challenging 21-cm observations are still some years away, but with the newly refurbished HST we can make immediate progress through a direct census of faint star-forming galaxies and studies of their stellar populations.

A first census of early galaxies

High-redshift galaxies are located by detecting the effect of intervening neutral hydrogen absorption on their colours. Even small amounts of neutral gas can extinguish the light from a galaxy at smaller wavelengths than the Ly α line in its rest frame, causing a ‘break’ in the observed galaxy spectrum. As a result, high-redshift galaxies are often referred to as Lyman-break galaxies^{38–41}, and the telltale signature of such a distant object is its disappearance or ‘drop-out’ from an imaging filter sensitive to the redshifted ultraviolet passband where the hydrogen absorption occurs. Identifying the highest-redshift galaxies and determining their influence on reionization therefore require an infrared-sensitive camera.

Excellent progress has been made using the drop-out technique to find high-redshift galaxies since May 2009, following the installation of the HST’s Wide Field Camera 3, a panoramic imager that includes a powerful infrared detector (WFC3/IR) operating in the wavelength range 850–1,700 nm. Taking into account the infrared field of view, pixel scale and efficiency, this instrument provides a 40-fold improvement in survey speed over the previous-generation instrument, the Near Infrared Camera and Multi-Object Spectrometer. The HST Wide Field Camera 3 undertook a series of deep images of the Hubble Ultra Deep Field (UDF; a 4.7-arcmin² area) reaching an optimal 5σ point source sensitivity of ~ 29 th magnitude in three broadband filters (Box 3). These data, together with shallower exposures in other areas, have provided the first convincing census of $z \approx 7$ galaxies and initial indications of galaxy populations at yet higher redshifts^{39–47}.

The most important result to emerge from the new WFC3/IR data has been the first robust determination of the volume density of galaxies of different luminosities at $z \approx 7$ (the luminosity function), based on over 50 sources seen so far in the various WFC3/IR campaigns. To these HST data sets can be added constraints based on $22z \approx 7$ candidates similarly detected to brighter limits (AB magnitude, $m_{\text{AB}} \approx 26$) using the NAOJ Subaru telescope⁴⁸ (Box 3), extending the range of the luminosity function. The depth of the HST images has been particularly advantageous, as the luminosity function increases steeply for intrinsically fainter sources, indicating that the bulk of the ultraviolet luminosity density from star-forming galaxies at $z \approx 7$ emerges from an abundant population of very faint systems.

BOX 1

The physics of reionization

The process of reionization follows the transition from the neutral IGM, at high redshift, to the ionized IGM we observe locally. The volume ionization fraction of the IGM, $Q_{\text{H II}}$, progresses from neutral ($Q_{\text{H II}} = 0$) to fully ionized ($Q_{\text{H II}} = 1$) according to a changing balance between the production rate of ionizing photons and the recombination rate^{13–16}. The production rate of IGM-ionizing photons

$$\frac{dn_{\text{ion}}}{dt} = f_{\text{esc}} \zeta_Q \rho_{\text{SFR}}$$

is the product of the co-moving star formation rate density, ρ_{SFR} (in units of $M_{\odot} \text{ yr}^{-1} \text{ Mpc}^{-3}$, where M_{\odot} is the solar mass), the number of hydrogen-ionizing photons per second per unit star formation rate, ζ_Q ($\text{s}^{-1} M_{\odot}^{-1} \text{ yr}$), and the fraction of photons that can escape a galaxy, f_{esc} . The rate dn_{ion}/dt is therefore fundamentally tied to the abundance and detailed astrophysics of early stars and galaxies. The recombination rate depends on the IGM temperature and the physical hydrogen density, which decreases with time according to the universal expansion factor, a^{-3} , and is enhanced in locally overdense regions by the ‘clumping factor’, $C_{\text{H II}} = \langle n_{\text{H}}^2 \rangle / \langle n_{\text{H}} \rangle^2$, where n_{H} is the hydrogen number density and angle brackets denote a cosmic average. Recent estimates suggest that $C_{\text{H II}} \approx 1$ – 6 (refs 16, 85). The recombination rate therefore depends primarily on atomic physics and the details of cosmological structure formation.

The observational requirements for determining whether galaxies had the dominant role in causing reionization are summarized as follows. First, ρ_{SFR} must be quantified, and is typically measured using the observable proxy of the rest-frame ultraviolet luminosity density, ρ_{UV} , through the luminosity distribution of high-redshift star-forming galaxies. Second, ζ_Q and f_{esc} must be observationally estimated by determining the relative effects of stars, dust and nebular emission on the rest-frame ultraviolet region of high-redshift galaxy spectral energy distributions or by direct observations in the Lyman continuum at lower redshift. To reionize the Universe fully, the integrated history of star formation must produce more than one ionizing photon per atom, such that ionizations exceed recombinations. For a standard stellar initial mass function, this requirement can be translated² into an effective co-moving stellar mass density of

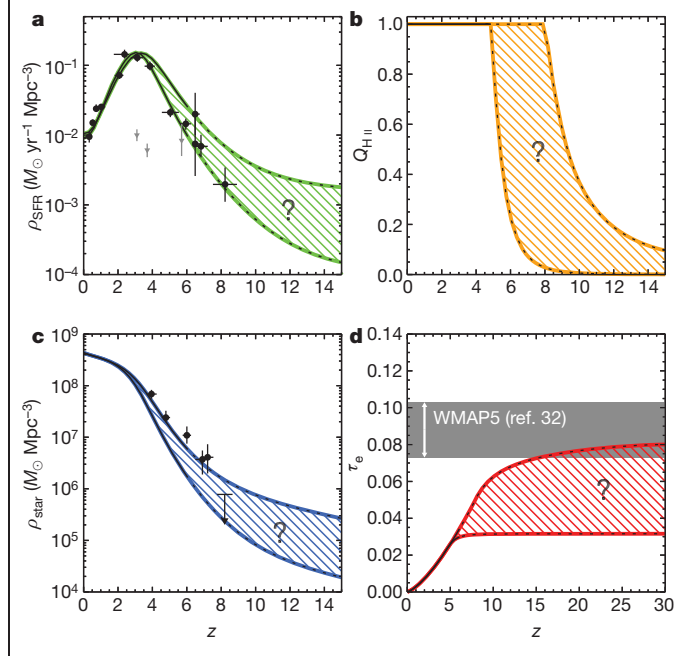
$$\rho_{\text{star}} \approx 2 \times 10^6 f_{\text{esc}}^{-1} M_{\odot} \text{ Mpc}^{-3}.$$

BOX 2

Observational probes of the reionization epoch

The observed rest-frame ultraviolet luminosity density of star-forming galaxies, which are the expected sources of the photons that cause reionization, provides the cosmic star formation rate density, ρ_{SFR} (Figure panel **a**: circles with 1-s.d. error bars^{40,41,62,86,87}; grey points indicate the contribution from Ly α -emitting galaxies⁶⁵). Also shown are illustrative model star formation histories for typical stellar ages of 10^8 yr, which are consistent with these observations (green area) and are based on an extension of the fitting form proposed in ref. 88. The models^{51,89} range from very metal-poor stars ($Z \approx 5 \times 10^{-6} Z_{\odot}$, where Z_{\odot} is the solar metallicity), at the upper boundary, to metal-rich stars ($Z \approx 2 Z_{\odot}$; see details below). We adopt the form $\rho_{\text{SFR}}(z) = [a + b(z/c)h]/[1 + (z/c)d] + g$ with $a = 0.009 M_{\odot} \text{ yr}^{-1} \text{ Mpc}^{-3}$, $b = 0.27 M_{\odot} \text{ yr}^{-1} \text{ Mpc}^{-3}$ and $h = 2.5$. In the metal-poor case we take $c = 3.7$, $d = 7.4$, $g = 10^{-3} M_{\odot} \text{ yr}^{-1} \text{ Mpc}^{-3}$, $f_{\text{esc}} = 0.3$ and $C_{\text{HII}} = 2$, and in the metal-rich case we take $c = 3.4$, $d = 8.3$, $g = 10^{-4} M_{\odot} \text{ yr}^{-1} \text{ Mpc}^{-3}$, $f_{\text{esc}} = 0.2$ and $C_{\text{HII}} = 6$.

The volume fraction of ionized hydrogen, Q_{HII} , implied by these models is shown in Figure panel **b** (orange region), where, consistent with the present data, the Universe becomes fully reionized ($Q_{\text{HII}} = 1$) at redshifts $z \approx 5.5$ – 8.5 . The observed stellar mass density (Figure panel **c**: data points with 1-s.d. error bars^{30,31,49–51,90}) also constrains the process of reionization because the stellar mass should trace the integral of the star formation rate density (blue shaded area) if most stars are long lived. Their relative agreement indicates that population III stars may not contribute significantly to the ultraviolet luminosity density at $z \approx 7$. Finally, the scattering optical depth, τ_e , of free electrons that polarize the CMB can also be measured³² (Figure panel **d**: grey shaded area). The model τ_e (red area) can be calculated from Q_{HII} by finding the path length through ionized hydrogen along the line of sight to the CMB. Producing the large electron scattering optical depth given the observed star formation rate density may be difficult without an evolving initial mass function, contribution from population III stars or yet-unobserved star formation at higher redshifts ($z > 10$).



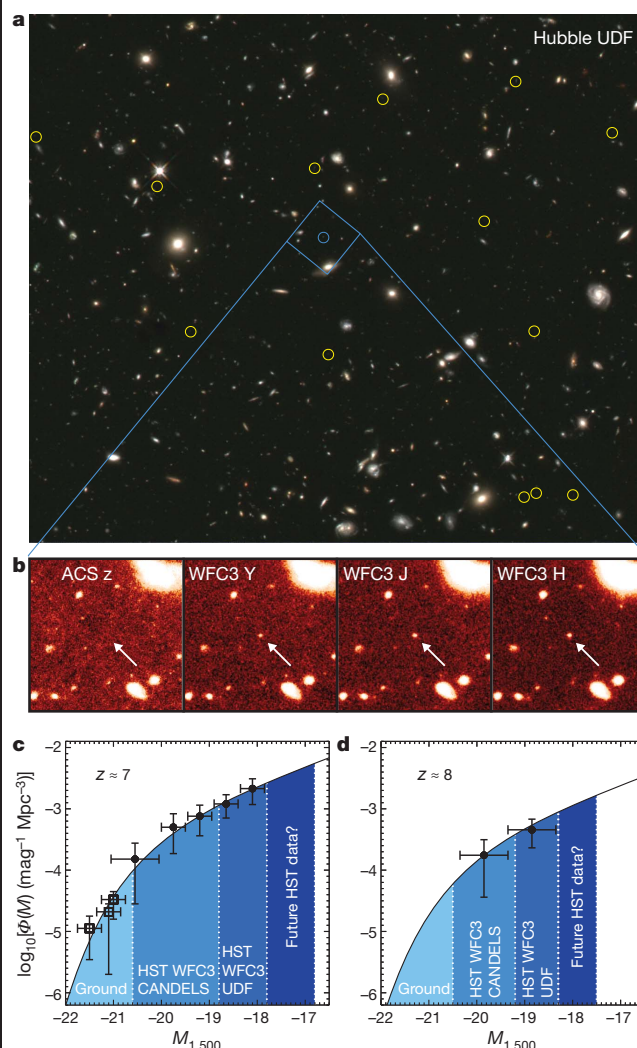
The most luminous $z \approx 7$ galaxies have also been detected individually by Spitzer^{49,50}, and many show strong continua at wavelengths greater than 400 nm in their rest frame, indicating established stellar populations

BOX 3

The discovery and study of galaxies at redshifts $z > 7$

The search for galaxies at extreme redshifts has been revolutionized by the successful installation of the new Wide Field Camera 3 in the HST. Figure panel **a** shows the deepest near-infrared image of the Hubble UDF yet taken with the new WFC3/IR camera, with the positions of newly discovered $z \approx 7$ – 8 galaxies^{39–41} indicated (circles). For a selected $z \approx 7$ candidate, Figure panel **b** shows that clear detections in the Y (1.05 μm), J (1.2 μm) and H (1.6 μm) bands with WFC3/IR, combined with the non-detection in the optical z band (0.85 μm), imaged with the Advanced Camera for Surveys (ACS), yield a secure photometric redshift³⁹ of $z = 7.2$.

The year-one WFC3/IR observations have already produced sufficient numbers of high-redshift galaxies to allow the first reliable determination³⁹ of the galaxy luminosity function, $\phi(M)$, at $z = 7$ (Figure panel **c**: circles with 1-s.d. error bars) and an initial estimate at $z = 8$ (Figure panel **d**: circles with 1-s.d. error bars). Here $M_{1,500}$ denotes the absolute AB magnitude at 1,500 Å in the galaxy rest frame. Large-area ground-based surveys⁴⁸ (Figure panel **c**: open squares with 1-s.d. error bars) are required to uncover the rarest luminous objects at the bright end of the luminosity function, and the newly approved HST Multi-Cycle Treasury Program, CANDELS, will provide a few hundred galaxies at luminosities around the characteristic luminosity L^* at $z = 7$ – 8 . More-sensitive imaging in the UDF is possible and is necessary to measure the number densities of fainter galaxies and to extend these studies to $z \approx 9$ – 10 .



of $\sim 10^9$ – 10^{10} solar masses. Stacking the Spitzer images of the more abundant, fainter population gives hints of a marginal signal corresponding to similar stellar populations whose mean mass is 10^8 – 10^9 solar masses⁵¹. The combination of HST and Spitzer has been very effective in addressing the key issue of identifying a sustained source of ionizing radiation. Although uncertainties remain, there is now reasonably convincing evidence that star formation in individual galaxies proceeded at a roughly constant rate over an extended period of 300 Myr, corresponding to the redshift range $7 < z < 10$.

The escape fraction of ionizing photons

In addition to counting the number of galaxies that produce energetic radiation, we must establish whether a sufficient fraction of the associated ultraviolet photons escape to allow reionization. To quantify the production rate of ionizing photons, dn_{ion}/dt , and conclusively determine the role of galaxies in cosmic reionization, we thus turn to the problem of determining the number, ζ_Q , of energetic Lyman-continuum photons per unit star formation rate produced by early stellar populations, and the fraction, f_{esc} , of such photons that freely escape each galaxy (Box 1). Although these quantities almost certainly vary significantly from one galaxy to the next, we can legitimately seek to establish a representative average for the purposes of determining the role of galaxies in reionization.

Because ultraviolet photons that have wavelengths below the Lyman limit (91.2 nm), and are therefore capable of reionizing the intergalactic atoms, are rapidly absorbed by neutral gas in the galaxy, the most direct way of estimating f_{esc} is to measure the emerging flux in this wavelength range. Such observations are intrinsically difficult because typically $f_{\text{esc}} \ll 1$ and the intervening IGM along the line of sight can absorb the escaping photons, further attenuating the detected Lyman-continuum flux. Despite these challenges, spectroscopic and narrow-band imaging observations^{52–55} have detected Lyman-continuum flux from galaxies at redshift $z \approx 3$, which is the practical redshift limit for this method. These measurements show that the effective escape fraction can vary widely from galaxy to galaxy, but imply characteristic values of $0.1 \lesssim f_{\text{esc}} \lesssim 0.2$ at $z \approx 3$.

The same experiment at redshift $z \approx 7$ is not technically feasible owing to the increased IGM absorption of high-redshift sources. However, another photometric signature of Lyman-limit photons as they migrate out of a young galaxy might be observable. If such a photon encounters neutral gas in the galaxy, it will probably ionize a portion of that gas and lead to line emission as well as free-free and bound-free scatterings between electrons and protons. These processes produce nebular radiation whose characteristic emission spectrum can be detected. Models incorporating both the stellar and nebular contributions to galactic emission^{56,57} predict a spectrum whose power-law slope, β (such that the flux scales with wavelength as $f(\lambda) \propto \lambda^\beta$), is strongly dependent on the number of escaping ionizing photons through ζ_Q and f_{esc} (ref. 42; Fig. 2). However, unlike the direct measurement of Lyman continuum photons at $z \approx 3$, this indirect method of estimating f_{esc} from the spectral character of $z \approx 7$ galaxies has yet to be conclusively demonstrated.

Lower-luminosity galaxies in the redshift range $4 < z < 7$ have steep ultraviolet slopes⁵⁸, $\beta < -2.0$, consistent with the hypothesis that these are relatively dust-free systems⁴⁴. Moreover, the new HST data has now indicated that this trend continues to higher redshift, where ultraviolet slopes with $\beta < -2.5$ have reportedly been observed^{42,44}. Because the youngest starburst galaxies in the local Universe have ultraviolet spectra with $\beta > -2.5$ (refs 58, 59) and extreme slopes ($\beta < -3$) may indicate population III stars^{60,61}, the large negative slope values derived from the new HST data are intriguing.

The ultraviolet slope measurements beyond $z \approx 7$ remain controversial, partly because of the photometric uncertainties involved. However, if verified with more-precise, deeper imaging, current models suggest that such steep slopes most likely indicate nearly dust-free, metal-poor stellar systems with large ζ_Q values or young galaxies with very little nebular spectra, as would be expected if the escape fraction was significant (for

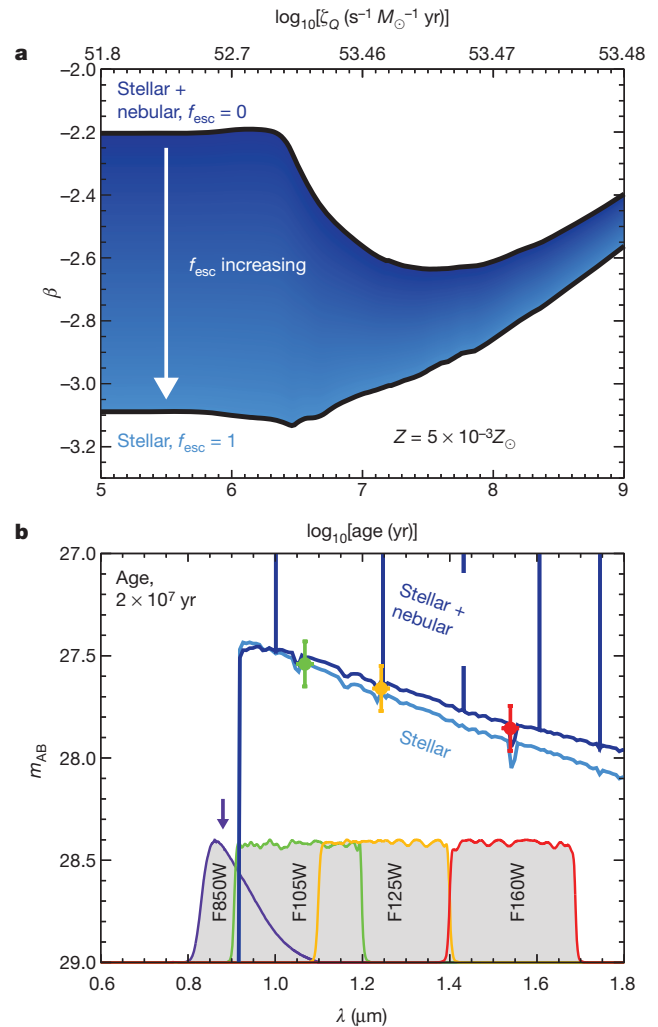


Figure 2 | Ionizing flux from high-redshift galaxies. The co-moving flux of hydrogen ionizing photons, dn_{ion}/dt , produced by galaxies depends on the total star formation rate density, ρ_{SFR} , the number of ionizing photons per unit star formation rate, ζ_Q , and the fraction, f_{esc} , of these photons that can escape galaxies to ionize the IGM. Most galaxies at $z \approx 7$ seem to be nearly dust free⁴², and the escape fraction may therefore reflect the internal ionization rate of gas within each galaxy. This internal ionization produces nebular emission⁷⁸ that can redden the spectra of nearly metal-free, star-forming galaxies. The colour of the galaxy determined using various filters (shaded areas in **b**) may therefore constrain ζ_Q and f_{esc} (ref. 57). **a**, Ultraviolet spectral slope, β , defined in terms of the flux density as $f(\lambda) \propto \lambda^\beta$, for the case of a metal-poor galaxy. We calculate β from stellar population synthesis models^{79,80} and our newly constructed model for the nebular spectrum^{78,81–84}. Galaxies with constant star formation rates and $f_{\text{esc}} \approx 1$ may appear extremely blue, whereas models with $f_{\text{esc}} \approx 0$ are redder owing to nebular emission. Measuring this slope for $z \approx 7$ galaxies is difficult. **b**, Model high-redshift galaxy spectra with and without nebular emission, along with the synthesized photometry in the available HST filters (data points). The typical UDF photometric uncertainty is ~ 0.25 mag per source and ~ 0.11 mag for stacked photometry of 20 objects (error bars, 1 s.d.). Hence, the current data are insufficiently deep to infer f_{esc} and ζ_Q unambiguously from the spectral slope.

example $f_{\text{esc}} > 0.2$). Reliably measuring the ultraviolet slopes for a population of $z \approx 7$ – 8 galaxies is within the capabilities of the HST if sufficiently deep exposures are used, and holds the key to constraining the desired combination of ζ_Q and f_{esc} values.

Our calculations (Fig. 3) find that the $z \approx 7$ photon budget, dn_{ion}/dt , for reionization is already balanced by the observed population of galaxies, although with some uncertainty. However, we show that ongoing, wider-area HST surveys, as well as future, deeper exposures, can resolve the remaining statistical uncertainties. Finalizing the connection between young galaxies and reionization will thus be possible well before the

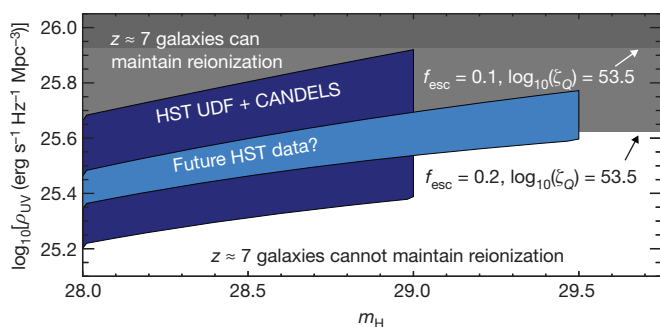


Figure 3 | Distant star-forming galaxies and reionization. Expected constraint on dn_{ion}/dt from the combined data of the UDF and the forthcoming CANDELS HST surveys, determined using a Fisher matrix calculation⁷³. Shown is the 3σ uncertainty in the rest-frame ultraviolet luminosity density, ρ_{UV} (the observable proxy for ρ_{SFR} ; dark blue region), and the improved constraint achieved by increasing the UDF limiting depth by 0.5 mag (light blue region). The light and dark grey regions show the ρ_{UV} ranges necessary for ionizations to balance recombinations in the IGM for $f_{\text{esc}} = 0.2$ and $f_{\text{esc}} = 0.1$, respectively, assuming that $\zeta_Q = 10^{53.5} \text{ s}^{-1} M_{\odot}^{-1} \text{ yr}$ (Fig. 2) and $C_{\text{HII}} \approx 2$. Increased C_{HII} , decreased f_{esc} and decreased ζ_Q require correspondingly greater ρ_{UV} for ionization–recombination balance. The putative escape fraction values are motivated by the steep observed ultraviolet slopes of $z \approx 7$ galaxies⁴² and detections of Lyman-continuum flux from $z \approx 3$ galaxies^{52–55}. These estimates are calculated using the measured luminosity function power-law faint-end slope³⁹ of $\alpha = -1.72$; a steeper slope would increase the inferred ultraviolet luminosity density and ionizing flux produced by galaxies. The depth of the UDF and the large area of CANDELS will constrain the abundance of $z \approx 7$ galaxies well enough to evaluate the hypothesis that galaxies trigger reionization, but a definitive test of this will require deeper data in the UDF with a commitment of additional exposure time comparable to the previous effort. With its new WFC3/IR instrument, the HST is capable of performing this experiment in the near future. m_{H} , AB magnitude in the H band.

arrival of the new facilities that will study this important reionization era in greater detail.

Charting the end of reionization

Just as quasars were originally used as distant beacons to probe the IGM at low redshift, the spectroscopy of distant galaxies can test for the presence of neutral hydrogen along the line of sight. A valuable spectral diagnostic in distant star-forming galaxies is the $\text{Ly}\alpha$ emission line at rest wavelength of 121.6 nm, which is produced internally by gas heated by young stars. The observability of $\text{Ly}\alpha$ emission is sensitive to the ionization state of the IGM, as the observed line strength can be attenuated by intervening neutral hydrogen. The challenge to using this technique to chart the end of reionization lies in isolating the impact of neutral hydrogen from other effects that may diminish the strength of the $\text{Ly}\alpha$ line, including dust.

The abundance of drop-out-selected galaxies decreases markedly from redshifts $z \approx 3$ to ~ 7 (ref. 62), yet the fraction of these galaxies showing intense $\text{Ly}\alpha$ emission increases with redshift⁶³. An important complementary technique for selecting high-redshift $\text{Ly}\alpha$ emitting galaxies⁶⁴ through narrowband filters yields a similar result⁶⁵ (Fig. 4a). As the strong line emitters are generally found to be galaxies with dust-free colours, the increasing fraction of line emitters suggests a reduced obscuration by dust at early times. We therefore expect most early star-forming galaxies to exhibit prominent $\text{Ly}\alpha$ emission, until we reach the era when the IGM became partly neutral, at which point the fraction of line-emitting galaxies should decrease.

Searches for a decrease in the fraction of $\text{Ly}\alpha$ -emitting galaxies could thus be a very effective probe of when reionization ended^{17,9,10}. Much excitement has been generated by a claimed drop in the abundance of $\text{Ly}\alpha$ emitters seen in narrowband imaging surveys with the Subaru telescope⁶⁶ in the short time interval (150 Myr) between $z = 5.7$ and 6.6. Further studies of a larger area⁶⁷ at $z = 6.6$ and a deeper narrowband search⁶⁸ at $z = 7$ also indicate a decreasing abundance of $\text{Ly}\alpha$ -emitting galaxies at $z > 6.5$. Such a change might highlight an increase in the IGM

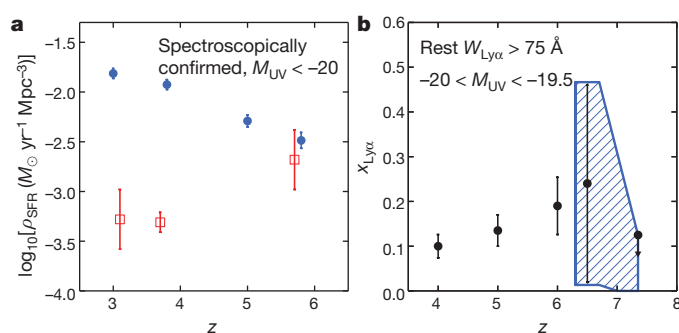


Figure 4 | Tracing the end of reionization with the $\text{Ly}\alpha$ line. **a**, Recent observations have revealed that the abundance of narrowband-selected $\text{Ly}\alpha$ -emitting galaxies (LAEs, red squares) is increasing⁶⁵ for $3 < z < 6$, when the IGM is known to be highly ionized, whereas the abundance of colour-selected Lyman-break galaxies (LBGs, blue points) is decreasing⁶². In the absence of changes in the IGM ionization state, we therefore expect most dust-free $z > 6$ galaxies to show powerful $\text{Ly}\alpha$ line emission. **b**, Observations at $z > 6$ have instead revealed a possible decrease in the prevalence of $\text{Ly}\alpha$ emitters⁶⁶ for $5.7 < z < 7.0$ (black points), as expected if reionization ended at $z \approx 6$. This trend may now be confirmed by spectroscopic surveys of $4 < z < 7$ LBGs that show a tentative decrease in the fraction of $\text{Ly}\alpha$ -emitting galaxies^{63,70,71} at $z > 6.3$ (blue hatched region). This result will require confirmation through infrared spectroscopy of $z > 7$ LBGs. Error bars, 1 s.d. M_{UV} , absolute magnitude in the ultraviolet; $W_{\text{Ly}\alpha}$, $\text{Ly}\alpha$ equivalent width; $x_{\text{Ly}\alpha}$, fraction of sample showing $\text{Ly}\alpha$ emission with $W_{\text{Ly}\alpha} > 75 \text{ \AA}$.

neutral fraction. The patchy character of the neutral IGM at the end of reionization should also affect the spatial clustering of $\text{Ly}\alpha$ emitters⁶⁹, but as no such signal has been detected the question of whether this drop arises from a change in IGM transmission remains unanswered.

To verify the potentially important decrease in the abundance of $\text{Ly}\alpha$ emitters beyond $z \approx 6$, the fraction of line-emitting galaxies needs to be determined at even earlier times. If the Subaru survey results arose from an increased neutral fraction, we would expect a continued drop in the fraction of line-emitting galaxies at higher redshift. A few $z \approx 7$ –8 gravitationally lensed Lyman-break candidates have been found using deep imaging in foreground massive clusters, but none has been spectroscopically confirmed^{70,71} and, thus, some candidates may be at lower redshift. However, intriguingly, no $\text{Ly}\alpha$ emission has been seen in any of these lensed sources, suggesting that there has been a continued increase in the neutral fraction (Fig. 4b). Recently, there has been a claimed spectroscopic detection of $\text{Ly}\alpha$ line emission from a $z = 8.6$ drop-out-selected galaxy⁷² found in the new HST data. If confirmed, the detection of a $\text{Ly}\alpha$ line emitter would be a milestone in the study of the state of the IGM during the reionization epoch. Extending the $\text{Ly}\alpha$ visibility test with other $z > 7$ candidates newly discovered by HST is therefore an important priority for which results can be expected soon.

Future prospects

Although the rapid advances in the study of the reionization epoch afforded by the new capabilities of the HST have been remarkable, a variety of outstanding observational issues remain. Accurate measures of the abundance of high-redshift galaxies and their luminosity distributions may still need to be improved, and a detailed study of the stellar populations in these galaxies is necessary to predict their output of ultraviolet photons robustly. Furthermore, the gaseous content of high-redshift galaxies and the structure of neutral hydrogen in the IGM external to these galaxies are largely unknown. Addressing these outstanding issues will require renewed effort at existing and forthcoming observatories over the next decade.

We argue that the HST can further exploit its new infrared capabilities to improve the census of high-redshift galaxies. The current limitations are the precision with which the faint-end slope, α , of the luminosity function is determined and the possibility of contamination from lower-redshift sources. We have carefully evaluated various survey strategies⁷³

and estimate that a now-approved, larger-area, multicycle HST survey, in combination with the deeper UDF exposures, will improve the bright end of the luminosity function and reduce the uncertainty in α to ± 0.15 . These observations will provide more-precise estimates of the overall ultraviolet luminosity and star formation rate densities. Further deep HST imaging resulting from an investment comparable to that made for the current UDF data would improve the uncertainty in α to ± 0.08 (Fig. 3). Such ultra-deep exposures, with a deployment of infrared filters carefully arranged to tighten constraints on the ultraviolet slopes of $z \approx 7$ galaxies, can also be used to identify higher-redshift galaxies reliably.

Detailed spectroscopic studies of the stellar populations of high-redshift galaxies are required to determine the abundance of young, hot stars and the importance of the gas and dust in absorbing photons that can reionize the IGM. Although comparable sources at lower redshift may soon be within reach of a new generation of infrared spectrographs nearing completion for 8–10-m-class telescopes, the NASA James Webb Space Telescope⁷⁴ (JWST) and the next generation of extremely large ground-based telescopes with apertures of 20–40 m (the Thirty Meter Telescope⁷⁵, the Giant Magellan Telescope⁷⁶ and the ESO European Extremely Large Telescope (<http://www.eso.org/public/teles-instr/e-elt.html>)) will ultimately resolve observational uncertainties about whether strong nebular emission pollutes some of the Spitzer photometry of $z \approx 7$ galaxies. If so, both the currently derived stellar masses and ages of high-redshift galaxies have been overestimated^{56,57}. Additionally, the JWST will use infrared cameras and spectrographs designed to probe the rest-frame ultraviolet and optical emission from galaxies beyond redshifts $z \approx 10$.

Forthcoming radio telescopes will give us information on the abundance and distribution of gas fuelling star formation in high-redshift galaxies, as well as the ionization state of hydrogen in the IGM. The Atacama Large Millimeter/submillimeter Array (<http://www.almaobservatory.org/>) may observe fine-structure carbon and nitrogen lines⁷⁷ in distant star-forming galaxies, allowing us to characterize how gas converts into the stars that produce ionizing photons. Future 21-cm interferometers (the Low Frequency Array (<http://www.lofar.org/>), the Murchison Widefield Array (<http://www.mwatelescope.org/>) and the Square Kilometre Array (<http://www.skatelescope.org/>)) have reionization studies as an important scientific justification, and will observe the topology of reionization through the spectral signatures of redshifted neutral hydrogen. These observatories will observe neutral hydrogen in the IGM ‘disappear’ as it is ionized during the reionization process, thereby providing an ‘inverse’ experiment to complement rest-frame ultraviolet observations with the HST and the JWST. As these observational facilities come online over the next decade, the reionization epoch will be thoroughly examined.

1. Penzias, A. A. & Wilson, R. W. A measurement of excess antenna temperature at 4080 Mc/s. *Astrophys. J.* **142**, 419–421 (1965).
2. Loeb, A. *How Did the First Stars and Galaxies Form?* (Princeton Univ. Press, 2010).
3. Gunn, J. E. & Peterson, B. A. On the density of neutral hydrogen in intergalactic space. *Astrophys. J.* **142**, 1633–1641 (1965).
4. Bromm, V., Yoshida, N., Hernquist, L. & McKee, C. F. The formation of the first stars and galaxies. *Nature* **459**, 49–54 (2009).
5. Gnedin, N. Y., Kravtsov, A. V. & Chen, H.-W. Escape of ionizing radiation from high-redshift galaxies. *Astrophys. J.* **672**, 765–775 (2008).
6. Razoumov, A. O. & Sommer-Larsen, J. Ionizing radiation from $z = 4$ –10 galaxies. *Astrophys. J.* **710**, 1239–1246 (2010).
7. Miralda-Escude, J. Reionization of the intergalactic medium and the damping wing of the Gunn-Peterson trough. *Astrophys. J.* **501**, 15–22 (1998).
8. Loeb, A. & Rybicki, G. B. Scattered Lyman alpha radiation around sources before cosmological reionisation. *Astrophys. J.* **524**, 527–535 (1999).
9. Santos, M. R. Probing reionization with Lyman α emission lines. *Mon. Not. R. Astron. Soc.* **349**, 1137–1152 (2004).
10. Malhotra, S. & Rhoads, J. E. Luminosity functions of Ly α emitters at redshifts $z = 6.5$ and $z = 5.7$: evidence against reionization at $z \lesssim 6.5$. *Astrophys. J.* **617**, L5–L8 (2004).
11. Dijkstra, M., Haiman, Z. & Spaans, M. Lyman alpha radiation from collapsing protogalaxies. I. Characteristics of the emergent spectrum. *Astrophys. J.* **649**, 14–36 (2006).
12. Zheng, Z., Cen, R., Trac, H. & Miralda-Escude, J. Radiative transfer modelling of Lyman alpha emitters. I. Statistics of spectra and luminosity. *Astrophys. J.* **716**, 574–598 (2010).
13. Madau, P., Haardt, F. & Rees, M. J. Radiative transfer in a clumpy universe. III. The nature of cosmological ionizing sources. *Astrophys. J.* **514**, 648–659 (1999).

14. Wyithe, J. S. B. & Loeb, A. Reionization of hydrogen and helium by early stars and quasars. *Astrophys. J.* **586**, 693–708 (2003).
15. Choudhury, T. R. & Ferrara, A. Experimental constraints on self-consistent reionization models. *Mon. Not. R. Astron. Soc.* **361**, 577–594 (2005).
16. Bolton, J. S. & Haehnelt, M. G. The observed reionization rate of the intergalactic medium and the ionizing emissivity at $z \gtrsim 5$: evidence for a photon-starved and extended epoch of reionization. *Mon. Not. R. Astron. Soc.* **382**, 325–341 (2007).
17. Miralda-Escude, J., Haehnelt, M. & Rees, M. J. Reionization of the inhomogeneous Universe. *Astrophys. J.* **530**, 1–16 (2000).
18. Furlanetto, S. R., Zaldarriaga, M. & Hernquist, L. The growth of H II regions during reionization. *Astrophys. J.* **613**, 1–15 (2005).
19. Gnedin, N. Y. & Fan, X. Cosmic reionization redux. *Astrophys. J.* **648**, 1–6 (2006).
20. Iliev, I. T. et al. Simulating cosmic reionization at large scales – I. The geometry of reionization. *Mon. Not. R. Astron. Soc.* **369**, 1625–1638 (2006).
21. McQuinn, M. et al. The morphology of H II regions during reionization. *Mon. Not. R. Astron. Soc.* **377**, 1043–1063 (2007).
22. Mesinger, A. & Furlanetto, S. Efficient simulations of early structure formation and reionization. *Astrophys. J.* **669**, 663–675 (2007).
23. Trac, H., Cen, R. & Loeb, A. Imprint of inhomogeneous hydrogen reionization on the temperature distribution of the intergalactic medium. *Astrophys. J.* **689**, L81–L84 (2008).
24. Aubert, D. & Teyssier, R. A radiative transfer scheme for cosmological reionization based on a local Eddington tensor. *Mon. Not. R. Astron. Soc.* **387**, 295–307 (2008).
25. Thomas, R. M. et al. Fast large-scale reionization simulations. *Mon. Not. R. Astron. Soc.* **393**, 32–48 (2009).
26. Finlator, K., Ozel, F. & Dave, R. A new moment method for continuum radiative transfer in cosmological reionization. *Mon. Not. R. Astron. Soc.* **393**, 1090–1106 (2009).
27. Choudhury, T. R., Haehnelt, M. G. & Regan, J. Inside-out or outside-in: the topology of reionization in the photon-starved regime suggested by Ly α forest data. *Mon. Not. R. Astron. Soc.* **394**, 960–977 (2009).
28. Alvarez, M. A., Busha, M., Abel, T. & Wechsler, R. Connecting reionization to the local universe. *Astrophys. J.* **703**, 167–171 (2009).
29. Springel, V. et al. Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature* **435**, 629–636 (2005).
30. Eyles, L. P. et al. The stellar mass density at $z \approx 6$ from Spitzer imaging of i' -drop galaxies. *Mon. Not. R. Astron. Soc.* **374**, 910–930 (2007).
31. Stark, D. P., Bunker, A. J., Ellis, R. S., Eyles, L. P. & Lacy, M. A new measurement of the stellar mass density at $z \sim 5$: implications for the sources of cosmic reionization. *Astrophys. J.* **659**, 84–97 (2007).
32. Komatsu, E. et al. Five-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological interpretation. *Astrophys. J. Suppl. Ser.* **180**, 330–376 (2009).
33. Fan, X. et al. Evolution of the ionizing background and the epoch of reionization from the spectra of $z \sim 6$ quasars. *Astron. J.* **123**, 1247–1257 (2002).
34. Fan, X., Carilli, C. L. & Keating, B. Observational constraints on cosmic reionization. *Annu. Rev. Astron. Astrophys.* **44**, 415–462 (2006).
35. Becker, G. D., Rauch, M. & Sargent, W. L. W. The evolution of optical depth in the Lyman alpha forest: evidence against reionization at $z \sim 6$. *Astrophys. J.* **662**, 72–93 (2007).
36. Kawai, N. et al. An optical spectrum of the afterglow of a γ -ray burst at a redshift of $z = 6.295$. *Nature* **440**, 184–186 (2006).
37. Furlanetto, S. R., Oh, S. P. & Briggs, F. H. Cosmology at low frequencies: the 21 cm transition and the high-redshift universe. *Phys. Rep.* **433**, 181–301 (2006).
38. Steidel, C. C., Pettini, M. & Hamilton, D. Lyman limit imaging of high-redshift galaxies. III. New observations of four QSO fields. *Astron. J.* **110**, 2519–2536 (1995).
39. McLure, R. J. et al. Galaxies at $z = 6$ –9 from the WFC3/IR imaging of the Hubble Ultra Deep Field. *Mon. Not. R. Astron. Soc.* **403**, 960–983 (2010).
40. Bouwens, R. J. et al. Discovery of $z \sim 8$ galaxies in the Hubble Ultra Deep Field from ultra-deep WFC3/IR observations. *Astrophys. J.* **709**, L133–L137 (2010).
41. Oesch, P. A. et al. $z \sim 7$ galaxies in the HUDF: first epoch WFC3/IR results. *Astrophys. J.* **709**, L16–L20 (2010).
42. Bouwens, R. J. et al. Very blue UV-continuum slope β of low luminosity $z \sim 7$ galaxies from WFC3/IR: evidence for extremely low metallicities? *Astrophys. J.* **708**, L69–L73 (2010b).
43. Oesch, P. A. et al. Structure and morphologies of $z \sim 7$ –8 galaxies from ultra-deep WFC3/IR imaging of the Hubble Ultra-Deep Field. *Astrophys. J.* **709**, L21–L25 (2010).
44. Bunker, A. et al. The contribution of high redshift galaxies to cosmic reionisation: new results from deep WFC3 imaging of the Hubble Ultra Deep Field. *Mon. Not. R. Astron. Soc.* (in the press); preprint at (<http://arxiv.org/abs/0909.2255>) (2009).
45. Yan, H. et al. Galaxy formation in the reionization epoch as hinted by Wide Field Camera 3 observations of the Hubble Ultra Deep Field. *Res. Astron. Astrophys.* **10**, 867 (2010).
46. Wilkins, S. M. et al. Probing L* Lyman-break galaxies at $z \sim 7$ in GOODS-South with WFC3 on HST. *Mon. Not. R. Astron. Soc.* **403**, 938–944 (2010).
47. Finkelstein, S. L. et al. On the stellar populations and evolution of star-forming galaxies at $6.3 < z < 8.6$. *Astrophys. J.* **719**, 1250–1273 (2010).

48. Ouchi, M. *et al.* Large area survey for $z \sim 7$ galaxies in SDF and GOODS-N: implications for galaxy formation and cosmic reionization. *Astrophys. J.* **706**, 1136–1151 (2009).
49. Gonzalez, V. *et al.* The stellar mass density and specific star formation rate of the universe at $z \sim 7$. *Astrophys. J.* **713**, 115–130 (2010).
50. Labbe, I., Bouwens, R. J., Illingworth, G. D. & Franx, M. Spitzer IRAC confirmation of z_{850} -dropout galaxies in the Hubble Ultra Deep Field: stellar masses and ages at $z \sim 7$. *Astrophys. J.* **649**, L67–L70 (2006).
This paper reports Spitzer infrared detections of the recent HST candidates at $z \approx 7$ indicating the likelihood of earlier star formation.
51. Labbe, I. *et al.* Ultradeep Infrared Array Camera observations of sub- L^* $z \sim 7$ and $z \sim 8$ galaxies in the Hubble Ultra Deep Field: the contribution of low-luminosity galaxies to the stellar mass density and reionization. *Astrophys. J.* **708**, L26–L31 (2010).
52. Bunker, A., Marleau, F. & Graham, J. R. Seeking the ultraviolet ionizing background at $z \sim 3$ with the Keck telescope. *Astron. J.* **116**, 2086–2093 (1998).
53. Steidel, C., Pettini, M. & Adelberger, K. Lyman-continuum emission from galaxies at $z \sim 3.4$. *Astrophys. J.* **546**, 665–671 (2001).
54. Shapley, A. *et al.* The direct detection of Lyman continuum emission from star-forming galaxies at $z \sim 3$. *Astrophys. J.* **651**, 688–703 (2006).
55. Iwata, I. *et al.* Detections of Lyman continuum from star-forming galaxies at $z \sim 3$ through Subaru/Suprime-Cam narrow-band imaging. *Astrophys. J.* **692**, 1287–1293 (2009).
56. Schaerer, D. & de Barros, S. The impact of nebular emission on the ages of $z \approx 6$ galaxies. *Astron. Astrophys.* **502**, 423–426 (2009).
57. Schaerer, D. & de Barros, S. On the physical properties of $z \sim 6$ –8 galaxies. *Astron. Astrophys.* **515**, 73–88 (2010).
58. Bouwens, R. J. *et al.* UV continuum slope and dust obscuration from $z \sim 6$ to $z \sim 2$: the star formation rate density at high redshift. *Astrophys. J.* **705**, 936–961 (2009).
59. Meurer, G. R., Heckman, T. M. & Calzetti, D. Dust absorption and the ultraviolet luminosity density at $z \sim 3$ as calibrated by local starburst galaxies. *Astrophys. J.* **521**, 64–80 (1999).
60. Bromm, V., Kudritzki, R. P. & Loeb, A. Generic spectrum and ionization efficiency of a heavy initial mass function for the first stars. *Astrophys. J.* **552**, 464–472 (2001).
61. Schaerer, D. The transition from population III to normal galaxies: Ly α and He II emission and the ionising properties of high redshift starburst galaxies. *Astron. Astrophys.* **397**, 527–538 (2003).
62. Bouwens, R. J., Illingworth, G. D., Franx, M. & Ford, H. UV luminosity functions at $z \sim 4, 5$, and 6 from the Hubble Ultra Deep Field and other deep Hubble Space Telescope ACS fields: evolution and star formation history. *Astrophys. J.* **670**, 928–958 (2007).
This is a comprehensive analysis of HST data indicating a decrease in the abundance of star-forming galaxies at $z > 4$.
63. Stark, D. P., Ellis, R. S., Chiu, K., Ouchi, M. & Bunker, A. Keck spectroscopy of faint $3 < z < 7$ Lyman break galaxies: - I. New constraints on cosmic reionisation from the luminosity and redshift-dependent fraction of Lyman-alpha emission. *Mon. Not. R. Astron. Soc.* (in the press); preprint at (<http://arxiv.org/abs/1003.5244>) (2010).
This paper discusses the visibility of Ly α line emission in galaxies as a tracer of the end of reionization.
64. Iye, M. *et al.* A galaxy at a redshift $z = 6.96$. *Nature* **443**, 186–188 (2006).
65. Ouchi, M. *et al.* The Subaru/XMM-Newton Deep Survey (SXDS). IV. Evolution of Lyman-alpha emitters from $z = 3.1$ to 5.7 in the 1 deg^2 field: luminosity functions and AGN. *Astrophys. J. Suppl. Ser.* **176**, 301–330 (2008).
66. Kashikawa, N. *et al.* The end of the reionization epoch probed by Lyman-alpha emitters at $z = 6.5$ in the Subaru Deep Field. *Astrophys. J.* **648**, 7–22 (2006).
This paper contains the first intriguing claim of a decrease in the abundance of Ly α emitters, possibly indicative of the end of the reionization epoch.
67. Ouchi, M. *et al.* Statistics of 207 Ly α emitters at a redshift near 7: constraints on reionization and galaxy formation models. *Astrophys. J.* (in the press); preprint at (<http://arxiv.org/abs/1007.2961>) (2010).
68. Ota, K. *et al.* Lyman alpha emitters at $z = 7$ in the Subaru/XMM-Newton Deep Survey Field: photometric candidates and luminosity function. *Astrophys. J.* **722**, 803 (2010).
69. McQuinn, M. *et al.* Studying reionization with Ly α emitters. *Mon. Not. R. Astron. Soc.* **381**, 75–96 (2007).
70. Richard, J. *et al.* A Hubble and Spitzer Space Telescope survey for gravitationally lensed galaxies: further evidence for a significant population of low-luminosity galaxies beyond $z = 7$. *Astrophys. J.* **685**, 705–724 (2008).
71. Kneib, J.-P., Ellis, R. S., Santos, M. R. & Richard, J. A probable $z \sim 7$ galaxy strongly lensed by the rich cluster A2218: exploring the dark ages. *Astrophys. J.* **607**, 697–703 (2004).
72. Lehnert, M. D. *et al.* Spectroscopic confirmation of a galaxy at redshift $z = 8.6$. *Nature* (in the press).
73. Robertson, B. E. Estimating luminosity function constraints from high-redshift galaxy surveys. *Astrophys. J.* **713**, 1266–1281 (2010).
This paper presents a statistical formalism for forecasting constraints from high-redshift galaxy surveys that incorporates uncertainty from cosmic variance.
74. Gardner, J. P. *et al.* The James Webb Space Telescope. *Space Sci. Rev.* **123**, 485–606 (2006).
75. TMT Science Advisory Committee. *Thirty Meter Telescope Detailed Science Case 2007* (<http://www.tmt.org/sites/default/files/TMT-DSC-2007-R1.pdf>) (TMT Observatory Corporation, 2007).
76. GMTO Corporation. *Giant Magellan Telescope Science Requirements* (http://www.gmto.org/sciencecase/GMT-ID-01405-GMT_Science_Requirements.pdf) (Giant Magellan Telescope Organization Corporation, 2006).
77. Walter, F. & Carilli, C. Detecting the most distant ($z > 7$) objects with ALMA. *Astrophys. Space Sci.* **313**, 313–316 (2008).
78. Osterbrock, D. E. & Ferland, G. J. *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei* 2nd edn, 67–91 (Univ. Sci. Books, 2006).
79. Bruzual, G. & Charlot, S. Stellar population synthesis at the resolution of 2003. *Mon. Not. R. Astron. Soc.* **344**, 1000–1028 (2003).
80. Chabrier, G. Galactic stellar and substellar initial mass function. *Publ. Astron. Soc. Pacif.* **115**, 763–795 (2003).
81. Karzas, W. J. & Latter, R. Electron radiative transitions in a Coulomb field. *Astrophys. J.* **6** (suppl.), 167–212 (1961).
82. Brown, R. L. & Mathews, W. G. Theoretical continuous spectra of gaseous nebulae. *Astrophys. J.* **160**, 939–946 (1970).
83. Sutherland, R. S. Accurate free-free Gaunt factors for astrophysical plasmas. *Mon. Not. R. Astron. Soc.* **300**, 321–330 (1998).
84. Anders, P., & Fritze-v. Alvensleben, U. Spectral and photometric evolution of young stellar populations: the impact of gaseous emission at various metallicities. *Astron. Astrophys.* **401**, 1063–1070 (2003).
85. Pawlik, A., Schaye, J. & van Scherpenzeel, E. Keeping the Universe ionised: photoheating and the high-redshift clumping factor of the intergalactic medium. *Mon. Not. R. Astron. Soc.* **394**, 1812–1824 (2009).
86. Schiminovich, D. *et al.* The GALEX-VVDS measurement of the evolution of the far-ultraviolet luminosity density and the cosmic star formation rate. *Astrophys. J.* **619**, L47–L50 (2005).
87. Reddy, N. A. & Steidel, C. C. A steep faint-end slope of the UV luminosity function at $z \sim 2$ –3: implications for the global stellar mass density and star formation rate in low-mass halos. *Astrophys. J.* **692**, 778–803 (2009).
88. Cole, S. *et al.* The 2dF galaxy redshift survey: near-infrared galaxy luminosity functions. *Mon. Not. R. Astron. Soc.* **326**, 255–273 (2001).
89. Salpeter, E. E. The luminosity function and stellar evolution. *Astrophys. J.* **121**, 161–167 (1955).
90. Stark, D. P. *et al.* The evolutionary history of Lyman break galaxies between redshift 4 and 6: observing successive generations of massive galaxies in formation. *Astrophys. J.* **697**, 1493–1511 (2009).

Acknowledgements We thank A. Klypin for the use of his cosmological simulation, and A. Loeb, A. Shapley and L. Hernquist for comments. B.E.R. acknowledges support from a Hubble Fellowship. R.S.E. acknowledges the hospitality of Leiden Observatory. J.S.D. acknowledges the support of the Royal Society through a Wolfson Research Merit award, and also the support of the European Research Council through the award of an Advanced Grant. R.J.M. acknowledges the support of the Royal Society through a University Research Fellowship. D.P.S. acknowledges financial support from an STFC postdoctoral research fellowship and a Schlumberger Research Fellowship at Darwin College.

Author Contributions B.E.R. and R.S.E. wrote the main manuscript text. B.E.R. performed the calculations presented in Box 2 and Figs 1–3. J.S.D. and R.J.M. prepared and analysed the data presented in Box 3. D.P.S. prepared and analysed the data presented in Fig. 4. All authors reviewed, discussed and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence should be addressed to B.E.R. (brant@astro.caltech.edu).

Quantum gravitational contributions to quantum electrodynamics

David J. Toms¹

Quantum electrodynamics describes the interactions of electrons and photons. Electric charge (the gauge coupling constant) is energy dependent, and there is a previous claim that charge is affected by gravity (described by general relativity) with the implication that the charge is reduced at high energies. However, that claim has been very controversial and the matter has not been settled. Here I report an analysis (free from the earlier controversies) demonstrating that quantum gravity corrections to quantum electrodynamics have a quadratic energy dependence that result in the electric charge vanishing at high energies, a result known as asymptotic freedom.

The standard model of particle physics is based in part on quantized Yang–Mills (or non-Abelian) gauge fields. The quantization of such fields leads to the important prediction that at high energy the strength of the gauge coupling constant governing the interaction of the Yang–Mills fields weakens, a phenomenon known as asymptotic freedom^{1,2}. This contrasts with a theory such as quantum electrodynamics (which describes electrons and photons), where the gauge coupling constant, the electric charge in this case, increases as the energy increases. A practical consequence of this is that for Yang–Mills theories that are asymptotically free, a perturbative approach based on a weak coupling constant becomes more reliable at high energy, whereas for quantum electrodynamics perturbation theory breaks down.

The key equations that govern the behaviour of the coupling constants in quantum field theory are the renormalization group Callan–Symanzik equations^{3,4}. If we let g denote a generic coupling constant, then the value of g at energy scale E , namely the running coupling constant, $g(E)$, is determined by

$$E \frac{dg(E)}{dE} = \beta(E, g) \quad (1)$$

where $\beta(E, g)$ is the renormalization group β -function. Asymptotic freedom is signalled by $g(E) \rightarrow 0$ as $E \rightarrow \infty$, requiring that $\beta < 0$ in this limit.

In the standard model of particle physics, gravity is usually ignored as it has an inessential role in most calculations of interest. Additionally, if we view Einstein's theory of gravity (general relativity) as a fundamental theory it has the undesirable property of non-renormalizability^{5–9}, and hence lacks predictive power. However, it is possible instead to view Einstein gravity as an effective theory that is only valid below some high energy scale^{10,11}. The cut-off scale is usually associated with the Planck scale, $E_P \approx 10^{19}$ GeV, which is the natural quantum scale for gravity. Above this energy scale, some theory of gravity other than Einstein's theory applies (string theory for example); below this energy scale, we can use quantized Einstein gravity to obtain reliable predictions^{10–12}. Adopting this effective field theory viewpoint means that it is perfectly reasonable to include Einstein gravity with the standard model of particle physics and to examine its possible consequences using quantum field theory methods.

With this in mind, a calculation was performed¹³ in Einstein–Yang–Mills theory that looked at the effects of quantum gravity on

the running Yang–Mills gauge coupling constant. This calculation found that quantum gravity leads to a correction to the renormalization group β -function (not present in the absence of gravity) that tends to result in asymptotic freedom; this holds even for theories (such as quantum electrodynamics) that are not asymptotically free when gravity is neglected. Possible phenomenological consequences of this result have been considered¹⁴.

The potential importance of the original calculation¹³ stimulated a number of further investigations that cast doubt on its findings. It was shown¹⁵ that a different choice of gauge condition led to the absence of any quantum gravity correction to the Yang–Mills β -function. Because of the possible dependence on gauge condition, a gauge-condition-independent version of the background field method^{16,17}, along with dimensional regularization¹⁸, was used¹⁹ and also found no quantum gravity contribution to the gauge coupling constant. The original result¹³ came from quadratic divergences that are automatically regulated to zero by dimensional regularization¹⁸.

The situation was analysed in a traditional Feynman diagram approach using standard Feynman rules, and it was shown²⁰ that if a momentum space cut-off was used the quadratic divergences could be made to cancel, leaving a result that was consistent with dimensional regularization. However, a novel method of regularization, called loop regularization, showed²¹ that the quadratic divergences were present, contradicting earlier results^{15,19,20}. Subsequent work has looked at the role of the cosmological constant^{22,23}, corrections in higher dimensions²⁴, and scalar^{25,26} and Yukawa fields²⁵. Additionally, related calculations have been performed in the exact renormalization group approach^{27–29}.

The issue of quadratic divergences and the possible gauge condition dependence is the central theme of the present Article. The generalized background field method^{16,17} that is used here differs from the usual one by the addition of an extra contribution that is essential for maintaining gauge condition invariance. The difference between the standard and generalized effective actions can be controlled by a single parameter, ν , that when set to one gives the gauge-condition-independent result and when set to zero gives the standard result, which coincides with a traditional Feynman diagram analysis. Gauge invariance is maintained by using a proper time^{30,31} cut-off. The importance of this choice of regularization is that, unlike many other methods, it is capable of revealing any possible quadratic divergences. I will show that the quadratic divergences that are obtained do not

¹School of Mathematics and Statistics, Newcastle University, Newcastle upon Tyne NE1 7RU, UK.

depend on ν and that the traditional background field method leads to a gauge-condition-dependent result for the charge renormalization. The contribution of the extra terms needed to obtain the gauge-condition-independent result is greatly simplified³² by choosing a particular gauge condition described below. Because the generalized effective action is independent of the choice made for the gauge condition, no generality is lost by this procedure. (A comprehensive review and justification of the method has been given recently³³.) I will show that, unexpectedly, the quadratic divergences do not cancel and the basic prediction originally found¹³ is correct, although with a slight modification of the β -function. Quantum electrodynamics is found to be asymptotically free if it is coupled to quantum gravity.

Effective action

The model studied is Einstein gravity with a cosmological constant coupled to quantum electrodynamics in four space-time dimensions. (The presence of the cosmological constant is not essential to the point that will be made, but is included for generality.) The basic fields are the space-time metric, $g_{\mu\nu}$, the electromagnetic field, A_μ , and the Dirac field, ψ . The principal aim is to calculate the quantum gravity contribution to the renormalization of the charge. To do this, it is sufficient to use the background field method, choose the background metric to be flat, choose the background Dirac field to vanish, choose the background electromagnetic field to correspond to a constant field strength, $\bar{F}_{\mu\nu}$, and concentrate on that part of the one-loop effective action that is divergent and involves the square of the background field strength.

A standard calculation shows that the effective action to one-loop order is given by

$$\Gamma^{(1)} = \frac{1}{2} \ln(\det \Delta^i_j) - \ln(\det Q_{\alpha\beta}) - \ln(\det(i\gamma^\mu \partial_\mu + e\gamma^\mu \bar{A}_\mu - im)) \quad (2)$$

The last term (where \bar{A}_μ is the background gauge field) is the result of performing a functional integral over the Dirac field. The middle term is the contribution from the ghost fields required to remove the unphysical degrees of freedom of the gravity and electromagnetic fields. The first term is the result of integrating over the space-time metric and electromagnetic fields; Δ^i_j is a second-order differential operator that can be found in earlier work^{22,23} and will not be written out here owing to its complexity. It is obtained by expanding the classical Einstein–Maxwell action about the background fields using $g_{\mu\nu} = \delta_{\mu\nu} + \kappa h_{\mu\nu}$ and $A_\mu = \bar{A}_\mu + a_\mu$. A Riemannian metric and standard conventions³⁴ for the Dirac fields are chosen, and $\kappa^2 = 32\pi G$, where G is the gravitational constant.

I use the gauge-condition-independent background field method^{16,17} as described above. The gauge conditions adopted are (with $h = \delta^{\mu\nu} h_{\mu\nu}$)

$$\chi_\lambda = \partial^\nu h_{\lambda\nu} - \frac{1}{2} \partial_\lambda h + \frac{\kappa}{2} \omega (\bar{A}_\lambda \partial^\mu a_\mu + \bar{F}_{\mu\lambda} a^\mu) \\ \chi = \partial^\mu a_\mu$$

To obtain the gauge-condition-independent result, it is essential to take $\omega = 1$ in our calculation. However, ω is kept general at this stage to indicate the gauge condition dependence of standard methods of calculation. The choice $\omega = 0$ is often made and is called the de Donder gauge.

The gauge conditions are incorporated in the usual way, by altering the action with a gauge-fixing term

$$S_{\text{GF}} = \frac{1}{\xi} \int d^4x \chi^\lambda \chi_\lambda + \frac{1}{2\zeta} \int d^4x \chi^2$$

Here ξ and ζ are two dimensionless parameters. The choice $\xi = 1 = \zeta$ is often made because it simplifies the calculation enormously. However, doing so makes it impossible to address the gauge condition dependence of the calculation. I therefore keep ξ and ζ general here,

noting that the gauge-condition-independent effective action is the same as that found by taking the limits $\xi \rightarrow 0$ and $\zeta \rightarrow 0$ (along with $\omega = \nu = 1$).

Heat kernel

For the operator Δ^i_j , the heat kernel, $K^i_j(x, x'; \tau)$, is defined by

$$-\frac{\partial}{\partial \tau} K^i_j(x, x'; \tau) = \Delta^i_k K^k_j(x, x'; \tau)$$

with boundary condition $K^i_j(x, x'; \tau = 0) = \delta^i_j \delta(x, x')$, where τ is the proper time^{30,31}. The Green's function, $G^i_j(x, x')$, for the operator Δ^i_j is

$$\Delta^i_k G^k_j(x, x') = \delta^i_j \delta(x, x') \quad (3)$$

It follows that the Green's function and the heat kernel are related by

$$G^i_j(x, x') = \int_0^\infty d\tau K^i_j(x, x'; \tau)$$

The importance of the heat kernel for quantum field theory arises from the existence^{31,35} of an asymptotic expansion as $\tau \rightarrow 0$:

$$K^i_j(x, x'; \tau) \approx (4\pi\tau)^{-n/2} \sum_{r=0}^\infty \tau^r E_r^i_j(x) \quad (4)$$

Here n is the space-time dimension (which I take to be four) and the heat kernel coefficients, $E_r^i_j(x)$, depend only locally on the details of the coefficients of the differential operator Δ^i_j . For many operators of interest in physics, explicit expressions for the first few coefficients are known^{31,35}; however, the operators needed for the present calculation are more general than those considered so far. (Some checks on our results follow by a different method³⁶.)

The divergent part of the effective action, as well as the Green's function, can be related to the heat kernel coefficients. Formally (before regularization)

$$L_A = \frac{1}{2} \ln(\det \Delta^i_j) = -\frac{1}{2} \int d^n x \int_0^\infty \frac{d\tau}{\tau} \text{Tr} K^i_j(x, x'; \tau) \quad (5)$$

The one-loop effective action (equation (2)) is then given by

$$\Gamma^{(1)} = L_A - 2L_Q - 2L_{\text{Dirac}} \quad (6)$$

As with the Green's function (equation (3)), the divergent part of equation (5) comes from the $\tau \approx 0$ limit of the proper time integral. To deal with this in a way that respects gauge invariance, general coordinate invariance and the structure of any quadratic divergences, a proper time cut-off, τ_c , is used: the lower limit of the proper time integration is replaced with τ_c . The divergences will appear³⁰ as $\tau_c \rightarrow 0$. Because of this, the divergent part of L_A , which we denote $\text{divp}(L_A)$, follows from using the asymptotic form of the heat kernel expansion (equation (4)).

To make contact with the standard renormalization group procedure, which uses an energy scale cut-off, I note that τ_c has units of squared length, or inverse squared energy, in units where the reduced Planck's constant and the speed of light equal one. The proper time cut-off can therefore be replaced with an energy cut-off, E_c , using $\tau_c = E_c^{-2}$. The divergent part of L_A becomes

$$\text{divp}(L_A) = -\frac{1}{32\pi^2} \int d^4x \left(\frac{1}{2} E_c^4 \text{Tr} E_0 + E_c^2 \text{Tr} E_1 + \text{Tr} E_2 \ln E_c^2 \right) \quad (7)$$

where $\text{Tr} E_0$, $\text{Tr} E_1$ and $\text{Tr} E_2$ are the traces of the first three coefficients appearing in equation (4). Previous work^{19,22,23,25,26} used dimensional regularization¹⁸ with the space-time dimension taken as $n = 4 + \varepsilon$ in the limit $\varepsilon \rightarrow 0$. In this case, the lower limit on the proper time integration can be kept as $\tau = 0$, and as $\varepsilon \rightarrow 0$ the divergent part of L_A contains a simple pole and is given by

$$\text{divp}(L_A) = \frac{1}{16\pi^2\varepsilon} \int d^4x \text{Tr} E_2 \quad (8)$$

Comparison of the dimensional regularization result (equation (8)) with that of the cut-off method (equation (7)) shows that the coefficient of the simple pole at $\varepsilon = 0$ in dimensional regularization is the same as the coefficient of $\ln E_c^{-1}$. The quartic divergence (proportional to E_c^4), and the quadratic divergence (proportional to E_c^2) do not appear in dimensional regularization; they are both regulated to zero. If the cut-off energy, E_c , is regarded as a fundamental scale in the effective field theory, then neglect of these terms could be significant. In considering the renormalization of the gauge coupling constant, it is possible to show, by calculating the E_0 coefficient, that there can be no contribution to the charge renormalization from the term involving E_c^4 in equation (7). However, there is a potential contribution from the quadratic divergence; in fact, I will demonstrate that such a divergence is present, as found in the original calculation¹³, and that it alters the result found using dimensional regularization.

It is clear from equation (7) that the central object of importance for deciding whether or not quadratic divergences are present lies in the expression $\text{Tr} E_1$. Although $\text{Tr} E_1$ is known for special operators^{31,35,36}, it is not known for the general operators that arise in the present calculation. The full details are technically involved and will be presented elsewhere. If dimensional regularization is chosen, with $n = 4 + \varepsilon$, then $G_j^i(x, x)$ has a simple pole³⁷ as $\varepsilon \rightarrow 0$ whose residue involves the heat kernel coefficient E_1 :

$$\text{divp}(G_j^i(x, x)) = -\frac{1}{8\pi^2\varepsilon} E_1^i{}_j(x)$$

It should be noted that dimensional regularization is used only as a technical device for calculating E_1 , and that this is distinct from any choice of regularization used for the effective action. The calculation of E_1 has been checked by another method, one that does not use dimensional regularization.

The heart of the calculation now involves finding the pole part of $G_j^i(x, x)$ and extracting from it the expression for E_1 . To accomplish this, I take the local momentum space approach³⁸, which uses a normal coordinate expansion of the operator \mathcal{A}_j^i . The general form of \mathcal{A}_j^i is

$$\mathcal{A}_j^i = (A^{\alpha\beta})^i{}_j \partial_\alpha \partial_\beta + (B^\alpha)^i{}_j \partial_\alpha + (C)^i{}_j \quad (9)$$

for coefficients $(A^{\alpha\beta})^i{}_j$, $(B^\alpha)^i{}_j$ and $(C)^i{}_j$ that depend on the space-time coordinates through the background field. Normal coordinates are introduced at x' , with $x^\mu = x'^\mu + y^\mu$, and all of the coefficients in equation (9) are expanded about $y^\mu = 0$. This gives

$$(A^{\alpha\beta})^i{}_j = (A_0^{\alpha\beta})^i{}_j + \sum_{n=1}^{\infty} (A^{\alpha\beta}_{\mu_1 \dots \mu_n})^i{}_j y^{\mu_1} \dots y^{\mu_n} \quad (10)$$

with similar expansions for $(B^\alpha)^i{}_j$ and $(C)^i{}_j$. The Green's function is Fourier expanded as usual

$$G_j^i(x, x') = \int \frac{d^n p}{(2\pi)^n} e^{ip \cdot y} G_j^i(p) \quad (11)$$

except that the Fourier coefficient, $G_j^i(p)$, can also have a dependence on the origin of the coordinate system x' that is not indicated explicitly.

By substituting equation (11) into equation (3) and using the expansion in equation (10), and those for $(B^\alpha)^i{}_j$ and $(C)^i{}_j$, it is possible to solve for $G_j^i(p)$ as an asymptotic series in $1/p$ beginning at order p^{-2} . If

$$G_j^i(p) = G_0^i{}_j(p) + G_1^i{}_j(p) + G_2^i{}_j(p) + \dots$$

where $G_r^i{}_j(p)$ is of order p^{-2-r} as $p \rightarrow \infty$, then to calculate the pole part of $G_j^i(x, x)$ as $n \rightarrow 4$ only terms up to and including $G_2^i{}_j(p)$ are needed. These can be found iteratively beginning with $G_0^i{}_j(p)$. The evaluation

of these terms is extremely complicated and most of the calculation was done using CADABRA³⁹.

The net result of this lengthy calculation is that the gravity and gauge field contributions yield

$$\begin{aligned} \text{Tr} E_1 = & \kappa^2 \left(\frac{3}{8} - \frac{3}{4} \omega + \frac{1}{8} \omega^2 + \frac{3}{8} \xi - \frac{1}{2} \omega \xi \right. \\ & \left. + \frac{3}{8} \omega \zeta - \frac{1}{32} \omega \xi \zeta + \frac{1}{32} \omega^2 \zeta \right) \bar{F}^2 \\ & + (12 + 8\xi^2 + 3\nu + \nu^2) A \end{aligned}$$

where A is the cosmological constant. One check on this result is that all terms which involve $1/\xi$ and arise at intermediate stages of the calculation cancel from the final result. Another useful check is that if the choice $\xi = 1 = \zeta$ is made then the result for E_1 follows from a standard result³⁵ and agrees with the result for $\text{Tr} E_1$ found above. It is noteworthy that ν , which parameterizes the difference between the standard and gauge-condition-independent effective actions, does not enter the term in the E_1 coefficient that multiplies the field strength \bar{F}^2 , and therefore cannot contribute to the quadratic divergences responsible for charge renormalization. This was noted earlier²³ using a different approach. (The parameter ν does occur in the term involving the cosmological constant, but this can have no role in charge renormalization.) To recover the result of the standard background field method, which must also agree with an analysis using Feynman diagrams and normal Feynman rules, we simply set $\nu = 0$. The coefficient of \bar{F}^2 in this term can be seen to depend explicitly on the two gauge-fixing parameters, ξ and ζ , as well as on the gauge condition parameter, ω . In this gauge-condition-dependent case, it still follows that there is a quadratic divergence, a result that is at variance with using a momentum space cut-off²⁰.

A similar procedure can be applied to the ghost field contribution (second term of equation (2)) and the Dirac field contribution (last term of equation (2)). The Dirac field has no quadratic divergence³⁴, because $\text{Tr} E_1 = 0$, but does have a logarithmic divergence coming from $\text{Tr} E_2$. The ghost field operator results in

$$\text{Tr} E_1 = -\frac{\kappa^2}{4} \omega^2 \bar{F}^2$$

Divergences and renormalization group

The overall result for the quadratically divergent part of the complete one-loop effective action (equation (6)) that involves \bar{F}^2 is

$$\begin{aligned} \Gamma_{\text{quad}}^{(1)} = & -\frac{\kappa^2 E_c^2}{32\pi^2} \left(\frac{3}{8} - \frac{3}{4} \omega + \frac{5}{8} \omega^2 + \frac{3}{8} \xi - \frac{1}{2} \omega \xi \right. \\ & \left. + \frac{3}{8} \omega \zeta - \frac{1}{32} \omega \xi \zeta + \frac{1}{32} \omega^2 \zeta \right) \int d^4x \bar{F}^2 \end{aligned}$$

This conclusively demonstrates the dependence of the standard result on both the gauge parameter and the gauge condition. If the limits $\xi \rightarrow 0$, $\zeta \rightarrow 0$ and $\omega \rightarrow 1$ are taken to obtain the gauge-condition-independent result discussed above, the non-zero result

$$\Gamma_{\text{quad}}^{(1)} = -\frac{\kappa^2 E_c^2}{128\pi^2} \int d^4x \bar{F}^2$$

is found.

This is not the complete divergent part of the effective action that involves \bar{F}^2 ; there is still the logarithmic contribution. It is possible to extend the method of calculation described above to calculate E_2 , but the results can be deduced from the known^{22,23} poles found for the first two terms of equation (6) and for the Dirac contribution³⁴. The net result for the divergent part of the effective action that involves \bar{F}^2 , and therefore contributes to charge renormalization, is

$$\text{divp}(\Gamma^{(1)}) = \left(-\frac{\kappa^2 E_c^2}{128\pi^2} - \frac{3\kappa^2 A}{256\pi^2} \ln E_c^2 + \frac{e^2}{48\pi^2} \ln E_c^2 \right) \int d^4x \bar{F}^2$$

From this, the renormalization group function that governs the running electric charge is calculated, using equation (1), to be

$$\beta(E, e) = \frac{e^3}{12\pi^2} - \frac{\kappa^2}{32\pi^2} \left(E^2 + \frac{3}{2} A \right) e \quad (12)$$

The first term on the right hand side of equation (12) is that present in the absence of gravity (found by letting $\kappa \rightarrow 0$) and results in the electric charge increasing with energy. The second term is the correction due to quantum gravity. For pure gravity with $A = 0$, or for a small value of A as suggested by present observational evidence⁴⁰, the quantum gravity contribution to the renormalization group β -function is negative and therefore tends to result in asymptotic freedom, in agreement with the original calculation¹³.

Outlook

Although the calculation has been done for quantum electrodynamics, similar conclusions follow for Yang–Mills fields with and without further matter present. It is possible that the running of the gauge coupling constants (found by solving equation (1) for a realistic gauge theory) can lower the unification scale relative to that expected in the absence of gravity¹³. In addition, if we allow space-time to have more than four dimensions, then the presence of extra dimensions could lower the effective gravitational scale to below the Planck scale⁴¹ and render the results of running charges measurable¹⁴ in CERN's Large Hadron Collider. There is also an intriguing connection with the weak-gravity conjecture⁴², which predicts that the natural cut-off should be a couple of orders of magnitude below the Planck scale. The weak-gravity conjecture translates into the requirement that the gravitational contribution to the renormalization group β -function should be less than that found in the absence of gravity⁴³.

Received 5 June; accepted 2 September 2010.

- Gross, D. & Wilczek, F. Ultraviolet behavior of non-Abelian gauge theories. *Phys. Rev. Lett.* **30**, 1343–1346 (1973).
- Politzer, D. Reliable perturbative results for strong interactions? *Phys. Rev. Lett.* **30**, 1346–1349 (1973).
- Callan, C. G. Broken scale invariance in scalar field theory. *Phys. Rev. D* **2**, 1541–1547 (1970).
- Symanzik, K. Small distance behaviour in field theory and power counting. *Commun. Math. Phys.* **18**, 227–246 (1970).
- 't Hooft, G. & Veltman, M. One-loop divergences in the theory of gravitation. *Ann. Inst. Henri Poincaré A* **20**, 69–94 (1974).
- Deser, S. & van Nieuwenhuizen, P. Nonrenormalizability of the quantized Einstein–Maxwell system. *Phys. Rev. Lett.* **32**, 245–247 (1974).
- Deser, S. & van Nieuwenhuizen, P. One-loop divergences of quantized Einstein–Maxwell fields. *Phys. Rev. D* **10**, 401–410 (1974).
- Deser, S. & van Nieuwenhuizen, P. Nonrenormalizability of the quantized Dirac–Einstein system. *Phys. Rev. D* **10**, 411–420 (1974).
- Deser, S., Tsao, H. & van Nieuwenhuizen, P. One-loop divergences of the Einstein–Yang–Mills system. *Phys. Rev. D* **10**, 3337–3342 (1974).
- Donoghue, J. F. Leading quantum correction to the Newtonian potential. *Phys. Rev. Lett.* **72**, 2996–2999 (1994).
- Donoghue, J. F. General relativity as an effective field theory: the leading quantum corrections. *Phys. Rev. D* **50**, 3874–3888 (1994).
- Burgess, C. P. Quantum gravity in everyday life: general relativity as an effective field theory. *Living Rev. Relativity* (<http://www.livingreviews.org/lrr-2004-5>) (2004).
- Robinson, S. P. & Wilczek, F. Gravitational correction to running of gauge couplings. *Phys. Rev. Lett.* **96**, 231601 (2006).
- Gogoladze, I. & Cheung, C. N. Probing the gravitational scale via running gauge couplings. *Phys. Lett. B* **645**, 451–454 (2007).
- Pietrykowski, A. R. Gauge dependence of gravitational correction to running of gauge couplings. *Phys. Rev. Lett.* **98**, 061801 (2007).
- Vilkovisky, G. A. The unique effective action in quantum field theory. *Nucl. Phys. B* **234**, 125–137 (1984).
- DeWitt, B. S. in *Quantum Field Theory and Quantum Statistics* Vol. 1 (eds Batalin, I. A., Isham, C. J. & Vilkovisky, G. A.) 191–222 (Hilger, 1987).
- 't Hooft, G. & Veltman, M. Regularization and renormalization of gauge fields. *Nucl. Phys. B* **44**, 189–213 (1972).
- Toms, D. J. Quantum gravity and charge renormalization. *Phys. Rev. D* **76**, 045015 (2007).
- Ebert, D., Plefka, J. & Rodigast, A. Absence of gravitational contributions to the running Yang–Mills coupling. *Phys. Lett. B* **660**, 579–582 (2008).
- Tang, Y. & Wu, Y.-L. Gravitational contributions to the running of gauge couplings. Preprint at (<http://arxiv.org/abs/0807.0331>) (2008).
- Toms, D. J. Cosmological constant and quantum gravitational corrections to the running fine structure constant. *Phys. Rev. Lett.* **101**, 131301 (2008).
- Toms, D. J. Quantum gravity, gauge coupling constants, and the cosmological constant. *Phys. Rev. D* **80**, 064040 (2009).
- Ebert, D., Plefka, J. & Rodigast, A. Gravitational contributions to the running Yang–Mills coupling in large extra-dimensional brane worlds. *J. High Energy Phys.* **02**, 028 (2009).
- Rodigast, A. & Schuster, T. Gravitational corrections to Yukawa and ϕ^4 interactions. *Phys. Rev. Lett.* **104**, 081301 (2010).
- Mackay, P. T. & Toms, D. J. Quantum gravity and scalar fields. *Phys. Lett. B* **684**, 251–255 (2010).
- Daum, J.-E., Harst, U. & Reuter, M. Running gauge coupling in asymptotically safe quantum gravity. *J. High Energy Phys.* **01**, 084 (2010).
- Shaposhnikov, M. & Wetterich, C. Asymptotic safety of gravity and the Higgs boson mass. *Phys. Lett. B* **683**, 196–200 (2010).
- Zanusso, O., Zambelli, L., Vacca, G. P. & Percacci, R. Gravitational corrections to Yukawa systems. *Phys. Lett. B* **689**, 90–94 (2010).
- Schwinger, J. S. On gauge invariance and vacuum polarization. *Phys. Rev.* **82**, 664–679 (1951).
- DeWitt, B. S. *The Dynamical Theory of Groups and Fields* (Gordon and Breach, 1965).
- Fradkin, E. S. & Tseytlin, A. A. On the new definition of off-shell effective action. *Nucl. Phys. B* **234**, 509–523 (1984).
- Parker, L. E. & Toms, D. J. *Quantum Field Theory in Curved Spacetime* (Cambridge Univ. Press, 2009).
- Toms, D. J. Background-field method and the renormalization of non-Abelian gauge theories in curved space-time. *Phys. Rev. D* **27**, 1803–1813 (1983).
- Gilkey, P. B. The spectral geometry of a Riemannian manifold. *J. Diff. Geom.* **10**, 601–618 (1975).
- Barvinsky, A. O. & Vilkovisky, G. A. The generalized Schwinger–DeWitt technique in gauge theories and quantum gravity. *Phys. Rep.* **119**, 1–74 (1985).
- Toms, D. J. Renormalization of interacting scalar field theories in curved space-time. *Phys. Rev. D* **26**, 2713–2729 (1982).
- Bunch, T. S. & Parker, L. Feynman propagator in curved spacetime: a momentum-space representation. *Phys. Rev. D* **20**, 2499–2510 (1979).
- Peeters, K. Cadabra: a field-theory motivated symbolic computer algebra system. *Comput. Phys. Commun.* **176**, 550–558 (2007).
- Spergel, D. N. et al. Wilkinson Microwave Anisotropy Probe (WMAP) three year results: implications for cosmology. *Astrophys. J. Suppl. Ser.* **170**, 377–408 (2007).
- Arkani-Hamed, N., Dimopoulos, S. & Dvali, G. R. The hierarchy problem and new dimensions at a millimeter. *Phys. Lett. B* **429**, 263–272 (1998).
- Arkani-Hamed, N., Motl, L., Nicolis, A. & Vafa, C. The string landscape, black holes and gravity as the weakest force. *J. High Energy Phys.* **06**, 060 (2007).
- Huang, Q. Gravitational correction and weak gravity conjecture. *J. High Energy Phys.* **03**, 053 (2007).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to the author (d.j.toms@newcastle.ac.uk).

Genomic and functional adaptation in surface ocean planktonic prokaryotes

Shibu Yooseph^{1*}, Kenneth H. Nealson^{1*}, Douglas B. Rusch¹, John P. McCrow¹, Christopher L. Dupont¹, Maria Kim¹, Justin Johnson¹, Robert Montgomery¹, Steve Ferriera¹, Karen Beeson¹, Shannon J. Williamson¹, Andrey Tovchigrechko¹, Andrew E. Allen¹, Lisa A. Zeigler¹, Granger Sutton¹, Eric Eisenstadt¹, Yu-Hui Rogers¹, Robert Friedman¹, Marvin Frazier¹ & J. Craig Venter¹

The understanding of marine microbial ecology and metabolism has been hampered by the paucity of sequenced reference genomes. To this end, we report the sequencing of 137 diverse marine isolates collected from around the world. We analysed these sequences, along with previously published marine prokaryotic genomes, in the context of marine metagenomic data, to gain insights into the ecology of the surface ocean prokaryotic picoplankton (0.1–3.0 μm size range). The results suggest that the sequenced genomes define two microbial groups: one composed of only a few taxa that are nearly always abundant in picoplanktonic communities, and the other consisting of many microbial taxa that are rarely abundant. The genomic content of the second group suggests that these microbes are capable of slow growth and survival in energy-limited environments, and rapid growth in energy-rich environments. By contrast, the abundant and cosmopolitan picoplanktonic prokaryotes for which there is genomic representation have smaller genomes, are probably capable of only slow growth and seem to be relatively unable to sense or rapidly acclimate to energy-rich conditions. Their genomic features also lead us to propose that one method used to avoid predation by viruses and/or bacterivores is by means of slow growth and the maintenance of low biomass.

Molecular taxonomy and phylogeny¹ revitalized the field of marine microbiology, allowing for the first time the realization that the ‘unseen’ and ‘unknown’ majority of uncultivated microbial taxa could be identified by their 16S ribosomal RNA genes, and identifying widespread clades of marine bacteria and archaea that had no cultivated representatives². More recently, metagenomics has revealed the extent of diversity not fully explained using the available genomes of cultivated marine microbes. When the *Sorcerer II* Global Ocean Sampling (GOS) expedition metagenomic data were published^{3,4}, it was remarkable in its duality of diversity. On the one hand, there seemed to be only a few taxonomic groups that appeared routinely in marine surface water samples. Of these groups, only three (*Pelagibacter*, *Prochlorococcus* and *Synechococcus*) were represented by cultivated marine microbes. On the other hand, despite their apparent ubiquity and abundance in the surface ocean, there was virtually no complete or near-complete genomic assembly of any of the cosmopolitan taxa, implying that these groups, as judged by 16S rRNA sequence, are internally extremely diverse. The GOS data set was also remarkable because of its apparent lack of relatedness to the entire collection of sequenced genomes: using a process called fragment recruitment, which is akin to *in silico* DNA hybridization, only a few genomes from the entire repertoire of sequenced genomes were found to have a significant number of GOS reads assigned to them.

Here we use a large collection of sequenced marine genomes and metagenomic data to show the presence of two major groups in the marine surface picoplankton with striking differences in their metabolic and physiological capabilities. One group, representing the abundant and cosmopolitan prokaryotic picoplankton, is characterized by small genome sizes and a gene content which suggests that these microbes are capable of only slow growth with little metabolic plasticity. The other group contains a variety of microbial taxa that are

found in the surface ocean in low abundance, waiting for nutritionally improved conditions. This group has a gene content that allows for the microbes to adapt to a feast-or-famine lifestyle, and thus occasionally reach high numbers and become a dominant biomass (that is, bloom). We posit that microbes in this group, when considered in an ecological context, are strongly influenced by the presence and/or activities of marine eukaryotes. Our findings also led to a testable hypothesis, which we call ‘cryptic escape’: a major strategy in the true marine picoplankton involves the maintenance of abundant ($\sim 10^5$ cells ml^{-1}) populations of very small cells (small genomes and low biomass), thus decreasing predation due to bacterivores (ciliates and flagellates) and perhaps bacteriophages.

Results

Overview of the data set and bioinformatics analyses

Our data set (Supplementary Material 1) consists of 197 marine genomes, 10.97 million GOS metagenomic reads and more than 45,000 16S rRNA sequences from GOS 16S PCR libraries and the Ribosomal Database Project⁵ (RDP). Of the marine genomes used in this study, 137 were sequenced, assembled and annotated by the J. Craig Venter Institute (JCVI) as part of the Marine Microbial Genome Sequencing Project (MMGSP; <http://www.moore.org/microgenome/>), which is funded by the Gordon and Betty Moore Foundation. The MMGSP is an international collaborative project that has a primary goal of obtaining genome sequences of ecologically relevant microbes from a variety of diverse marine environments around the world. The MMGSP genomes are a resource for metagenome interpretation and provide insights into the metabolic repertoire and diversity of the marine microbial ecosystem. Many of these microbes were isolated from surface or near-surface waters. The MMGSP genomes thus theoretically complement the metagenomic data gathered by the GOS expedition.

¹J. Craig Venter Institute, Rockville, Maryland 20850, USA.

*These authors contributed equally to this work.

In the GOS data collection, aquatic microbes were size-fractionated by serial filtration through 20.0- μm , 3.0- μm , 0.8- μm and 0.1- μm filters. The microbes in the 0.1–3.0- μm size range are collectively referred to here as the marine surface picoplankton. The GOS data used in the present analysis are primarily from the 0.1- μm group (92% of the reads); the remainder consists of representation from the 0.8- μm (5% of the reads) and 3.0- μm (3% of the reads) groups. These were collected from the northwest Atlantic/tropical Pacific transect³ and the later Indian Ocean transect of the *Sorcerer II* GOS expedition.

We assessed the abundance of the sequenced genomes in the picoplankton from their representation in both metagenomic and 16S rRNA PCR libraries that are part of the GOS data set. We used GOS 16S rRNA sequences to assess the relative abundance and geographical distribution of marine taxa that are as yet uncultivated but whose sequencing could explain a larger portion of the GOS data. The 197 sequenced marine genomes were grouped into high- and low-abundance classes on the basis of their occurrence in the GOS metagenomic data. We identified differences in gene content and protein functional groups (pathways and modules) between these classes, and used this to characterize properties associated with the abundant and widespread marine surface picoplankton.

Overlap with GOS data at protein family level

The 137 MMGSP genomes together constitute 552 megabases of DNA sequence and 526,366 proteins (Supplementary Table 1). Various phyla are represented by these genomes, with the most highly represented being Proteobacteria (63.5%), Cyanobacteria (12.4%) and Bacteroidetes (11.6%). These microbes embody a range of physiological diversity and include carbon fixers, photoautotrophs, photoheterotrophs, nitrifiers and methanotrophs. To assess the diversity and representation of proteins in the GOS data, we clustered the 526,366 proteins together with a comprehensive set of proteins from publicly available genomic and metagenomic data sets^{6,7}. We found that the MMGSP genomes have high overlap with the GOS data at the protein family level: 78% of the MMGSP proteins fall into clusters that contain GOS sequences and these clusters account for 82.3% of the protein predictions based on GOS reads. On the basis of the clustering, on average 15.4% of the proteins in an MMGSP genome are orphans⁸, that is, they do not show sequence similarity to any known proteins. These genomes provide valuable context to inferences made using metagenomic data: 12.4% of previously GOS-only protein families⁷ contain MMGSP proteins. We had previously predicted⁷ that certain protein domains that were thought to be kingdom specific had examples in other kingdoms; these predictions were verified using the MMGSP genomes. The domains include some previously thought to be specific to eukaryotes (for example the indoleamine 2,3-dioxygenase domain (Pfam ID, PF01231) and the MAM domain (Pfam ID, PF00629)), and a domain previously thought to be specific to archaea (an HTH DNA-binding domain (Pfam ID, PF04967)).

Recruitment of metagenomic reads to sequenced genomes

We used a fragment recruitment tool³ to assign GOS reads to sequenced genomes, where a recruited read was assigned to a single best-matching genome. Although it is not a phylogeny-based method, fragment recruitment nevertheless helps to assess the representation of a given reference genome or its taxonomic neighbours in a given sample of reads. Of the 10.97 million GOS reads used in this study, 24.5% recruited to the 197 sequenced marine genomes with nucleotide identity matches of $\geq 50\%$. The distribution of the recruitment of reads to the genomes is skewed: although the ten most highly recruiting genomes account for 84% of the recruited reads, most of the genomes account for only very small proportions of the recruited reads (Fig. 1). A separate collection of 740 non-marine genomes recruited only 1.5% of the GOS reads.

Recruitment quality was assessed using two measures: depth of coverage, which is defined as the average number of reads covering

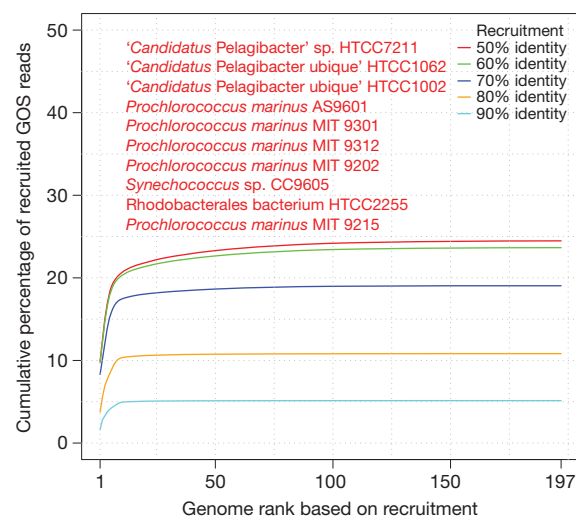


Figure 1 | Fragment recruitment of GOS reads to the 197 sequenced genomes at different nucleotide identity match thresholds. The x axis shows genome rank (from highest to lowest) based on the number of GOS reads recruited. The y axis shows the cumulative percentage (relative to the total number of GOS reads) of those reads that are recruited. Most of the sequenced genomes are part of the long tail of the distribution. Of the recruited reads, the majority are recruited to a few genomes, with the ten most highly recruiting genomes accounting for 84%, 85%, 91%, 95% and 96% of recruited reads for identity match thresholds of 50%, 60%, 70%, 80% and 90%, respectively. The ten most highly recruiting genomes at the 50% identity match threshold are listed.

a base pair in the reference genome, and fraction of the genome covered by the reads. At the 50% identity threshold, 17% of the genomes have a depth of coverage of ≥ 1 and in 46% of the genomes at least 10% of the genome is covered by the GOS reads (Supplementary Table 2). Only a few organisms in the cultivated set qualify as major constituents in the surface ocean picoplanktonic communities (Supplementary Fig. 1). Most of these are, not surprisingly, in the groups *Pelagibacter*, *Prochlorococcus* and *Synechococcus*, and of these some have very high depths of coverage. A few others are alpha-, beta- or gammaproteobacteria or flavobacteria. A crenarchaeota genome (*Nitrosopumilus maritimus* SCM1) also has a high depth of genome coverage. Overall, however, most cultivated microbes, including those from groups such as *Shewanella* and *Vibrio*, for which there are several sequenced members, are rarely abundant in the surface ocean (Supplementary Material 2).

Genomes from isolates within the same taxonomic clade recruit metagenomic reads in a pattern consistent with the geography of the sampling locations. The '*Candidatus Pelagibacter*' sp. HTCC7211 strain, which was isolated from the Sargasso Sea, recruits the largest number of GOS reads, in comparison with the cold-water strains HTCC1062 and HTCC1002, which were isolated from the Oregon coast of the Pacific Ocean. The relative recruitment-based ranking of the sequenced isolates changes as we go from the 0.1- μm -filter data to the 3.0- μm -filter data. Whereas the *Pelagibacter* group is dominant in the 0.1- μm -filter data, the *Prochlorococcus* and *Synechococcus* groups become the dominant recruiting groups in the data from the larger filters (Supplementary Material 2). The recruitment patterns for the sequenced isolates across the different GOS sampling sites (Fig. 2) are in agreement with recruitment data³ from when only a few genomes of each group were available. Of the taxonomic groups that recruit the most, the *Pelagibacter* group is generally present in all GOS samples whereas the *Prochlorococcus* group is notably absent⁹ in cold-water GOS samples.

16S rRNA analysis reveals the remaining uncultured majority

Fewer than 25% of the GOS metagenomic reads were recruited to the 197 marine genomes, posing the question of how well the sequenced

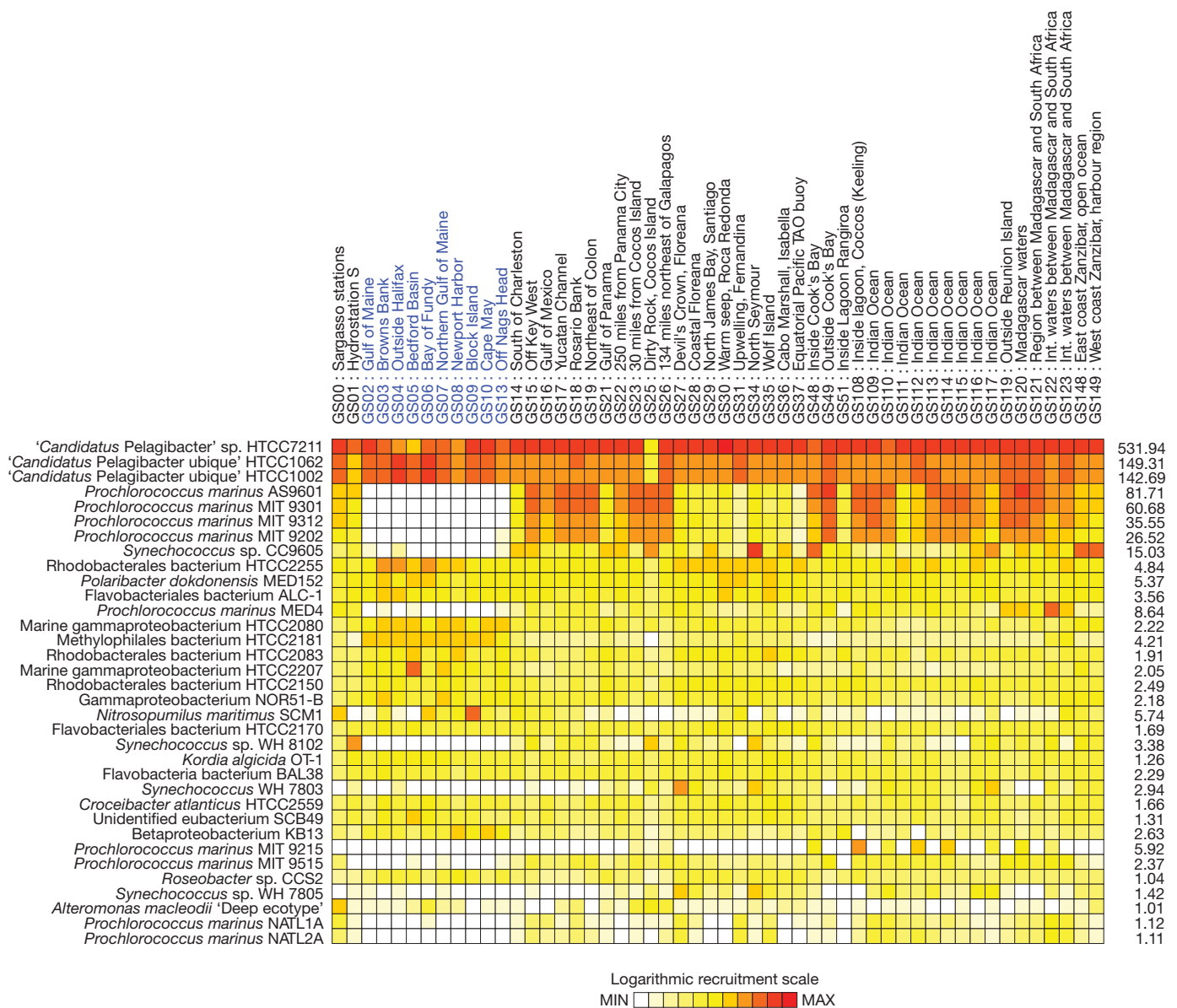


Figure 2 | Abundance and distribution of sequenced marine genomes in different GOS samples, based on fragment recruitment. The raw number of reads from a GOS site recruiting to a genome was normalized assuming 100,000 total reads for each GOS site. Logarithms of these numbers were subsequently taken and each value was assigned a colour relative to the maximum value seen

marine genomes represent the 16S-rRNA-based taxonomic space of surface marine bacterioplankton. To determine this, we compared their 16S rRNA genes with 37,860 GOS 16S PCR library sequences plus 8,471 marine bacterial 16S sequences obtained from the RDP⁵. We used a 97% identity match as the cut-off for identification of a bacterial species or operational taxonomic unit (OTU)^{10,11}. By this criterion, 15,642 sequences (33.7%) are recruited to 16S sequences of sequenced marine genomes; 16S rRNA genes from 740 non-marine genomes recruit an additional 709 sequences (1.5%). As was observed with fragment recruitment, the recruitment of 16S rRNA sequences from surface marine samples is highly skewed, with most of the 16S sequences recruiting to only a few genomes (Fig. 3).

We determined the phylogenetic distribution of the uncultured OTUs in 16S PCR samples (Methods and Supplementary Material 3). The abundant OTUs belong to bacterial classes that include alpha-, beta-, gamma- and deltaproteobacteria, actinobacteria, flavobacteria, sphingobacteria and cyanobacteria (Supplementary Fig. 2). There are many

(MAX = 17,623 reads): from light to dark, the colours represent 0 (MIN) to MAX in increments of MAX/10. Normalized values of <100 reads (or 0.01%) are set to 0. The depth of coverage of each genome is shown at the end of its corresponding row. Cold-water GOS samples are highlighted using blue text.

abundant, but as yet unsequenced, OTUs that are phylogenetically proximal to their sequenced counterparts in the groups *Pelagibacter*, *Prochlorococcus* and *Synechococcus*. The phylogeny also reveals the presence of abundant gammaproteobacteria SAR86¹² (and SAR86-related) and actinobacteria OTUs for which there are no cultured and sequenced representatives as yet. To determine the geographic distribution of the abundant OTUs, we scored the presence or absence of their constituent sequences across 35 GOS sampling sites. Although many of the abundant OTUs are geographically widely distributed in the oceans, several OTUs are cold- or warm-water specific (Fig. 4). The abundant OTUs from this analysis are candidates for isolation and/or sequencing (for instance by single-cell genomics techniques), the better to understand the marine picoplankton and provide more recruitment to the GOS data than we have seen so far. An analysis of 16S sequences from the GOS metagenomic data supports the findings, based on 16S PCR data, on the representation of the sequenced genomes and the abundant uncultured OTUs (Supplementary Material 1 and Supplementary Material 3).

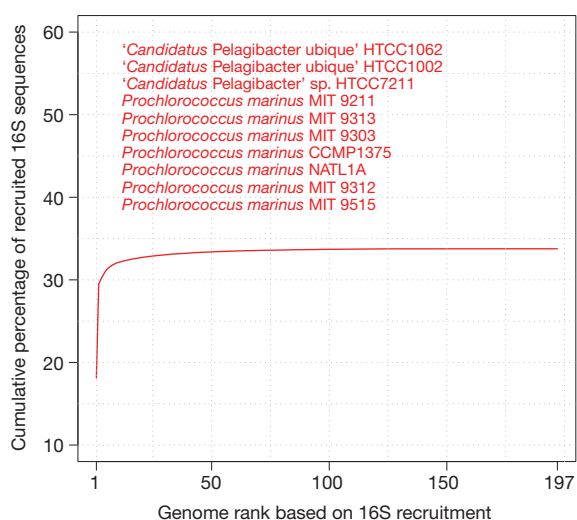


Figure 3 | Recruitment of 16S sequences to the 197 sequenced genomes. The x axis shows genome rank based on the number of 16S sequences recruited at the 97% identity match threshold. The y axis shows the cumulative fraction (relative to the 48K sequences) of recruited 16S sequences. The ten most highly recruiting genomes at the 97% identity match threshold are listed.

Genomic and functional characterization

To characterize the metabolic and physiological capabilities of the surface picoplankton, we examined the distribution of various genomic features (guanine–cytosine content (%GC) and genome size) and protein families/pathways in the sequenced marine genomes in relation to the depth of coverage of these genomes (Methods, Supplementary Material 4 and Supplementary Fig. 3). We grouped the 197 marine genomes into two sets: the high-recruiting genomes (HRGs), consisting of 34 genomes with coverage ≥ 1 , and the low-recruiting genomes (LRGs), consisting of 163 genomes with coverage < 1 . The genomes in the HRG set tend to have smaller size and lower %GC.

We identified protein groups that are differentially distributed between the HRG and LRG sets, even when genome size is taken into account. The slope of the trend line in each category was noted, with a positive slope indicating an overabundance in HRGs for that category and a negative slope indicating an underabundance. All the differentially distributed protein categories (Supplementary Table 3) were organized into higher-order functional groups, which included transcriptional regulation, transport, metabolism, biosynthesis, motility, chemotaxis, secretion, degradation, photosynthesis, and repair and replication.

Our analysis reveals distinct physiological differentiation and offers insights into the functional capabilities of the HRG and LRG sets. One of the most noticeable differences between the two sets is that genomes in the HRG set are characterized by the lack of many functional and regulatory genes. Transcriptional regulation is under-represented, and genes for energy-linked uptake (sugars and amino acids) and efflux (cations, drugs and so on) are nearly absent. Protein secretion to the exterior (for example extracellular proteases, DNases and chitinases) is curtailed, although transport across the cellular membrane to the periplasm seems to be an abundant character. In addition, motility and chemotaxis are absent, and genes known to be involved in quorum sensing (*luxI*, *luxR* and *luxS*), which usually deal with large populations of microbes and/or biofilm populations, are notably absent from the HRG set. Genes coding for functions involved in anaerobic metabolism are greatly decreased in number, whereas genes for aerobic metabolism, intermediary metabolism (glycolysis and gluconeogenesis, the tricarboxylic acid cycle and the pentose phosphate pathway), biosynthesis (amino acids, fatty acids, haem and vitamins), RNA synthesis, and repair and replication seem to be intact.

Transcriptional regulation of nearly every type is greatly decreased in the HRG set, and sensory domains of many types (for example histidine

kinases, the GGDEF domain and the methyl-accepting chemotaxis protein domain) are absent or greatly diminished. Also, energy-linked transport systems (for example proton-linked antiporters, PTS-sugar uptake and ATPase-linked efflux systems) are nearly absent from the HRG set, as is transport of amino acids and inorganic ions using ATPase-linked transporters. Similarly, efflux systems for toxic metals and drugs are notably absent.

Another curiosity with regard to transport is the apparent absence from the HRG set of systems for the uptake or synthesis of siderophore-like iron chelates. Because of the small genome size, the genomes in the HRG set have very few iron-requiring proteins and minimal iron requirements¹³. Despite this, iron will still be required for a few fundamental pathways, such as respiration in the heterotrophs, photosynthesis in the autotrophs and the GS–GOGAT cycle in both. On the basis of the genomic characteristics of the HRG set, it seems likely that they take up unchelated Fe(III) or Fe(II). Calculations suggest that this pool should support partial growth of cyanobacteria in even the most iron-depleted oceanic regions¹⁴. Furthermore, one of the organisms in the HRG set, *Synechococcus* WH8102, may be able to reduce Fe(III)–siderophore complexes, thereby increasing the pool of unchelated iron for uptake¹⁵, although the molecular identity of the reductase is unknown.

The secretion of proteins to the cell exterior is another function that is not well represented in the HRG set. This includes proteins such as serine proteases, metalloproteases, chitinases and DNases (involved in the metabolism of insoluble substrates), as well as pilin and flagellin (self-assembling proteins involved in cell attachment and/or motility). Secretion of proteins to the exterior through the general secretory pathways is virtually absent, suggesting that these microbes are not readily adaptable to the use of insoluble substrates or to a surface-associated lifestyle. However, protein excretion to the periplasm has been not only preserved in the HRG set, but is enriched, suggesting that periplasmic proteins may have an important role in the lives of the marine picoplankton. For example, the twin arginine transport genes (*tat*) and the secretion genes *secA*, *secB* and *secC*, all of which are involved in movement of proteins from the cytoplasm to the periplasm, are well represented in the HRG set.

Perhaps not surprisingly, the genes involved in motility and chemotaxis are strongly under-represented in the HRG set. This includes the structural genes mentioned above (pilin and flagellin synthesis), as well as a large array of sensory genes involved in sensing specific compounds and responding to these compounds through motility and chemotaxis. A number of genes involved in anaerobic metabolism are also notably under-represented in the HRG set. These include genes for low-level oxygen respiration (cytochrome-*bd* complex), genes for nitrate reduction, the anaerobic sigma factor (σ^{54}), regulatory genes involved in aerobic/anaerobic adaptation (such as *arcBA*) and genes for molybdate uptake and processing.

Clustered regularly interspaced short palindromic repeat (CRISPR) arrays and Cas genes, which form a system involved in bacterial defence against phages, are under-represented in the HRG set, as demonstrated by the counts of CRISPR arrays and genes plotted versus depth of coverage (Supplementary Fig. 3).

Most of the differentially distributed protein categories can be labelled as (–/–), that is, they show negative trends in both the original and the genome-size-normalized data (Supplementary Table 3). Essentially, many protein functions have been lost from the HRG set. However, some protein categories are preferentially retained, as shown by the trends of (–/+) or (+/+); none were (+/–). The positive trends in the normalized data for some categories were the result of a process that is seen in a subset of the genomes in the HRG set (and that is absent from the rest). Examples include protein categories such as carboxysomes and photosystems, which are part of the photosynthesis machinery present in the photoautotrophs in the HRG set. The positive trends seen in other categories, such as the various biosynthesis pathways, reflect core processes that are essential for all of



Figure 4 | Distribution of the 50 largest uncultured OTUs in GOS 16S PCR samples. Red and white squares respectively indicate the presence in and absence from a GOS sampling site of the corresponding OTU sequence. The number at the end of each row denotes the number of GOS sequences in that OTU. Cold-water GOS samples are highlighted using blue text.

these genomes. Other positive trends provide insight to the interaction of these organisms with the environment. For instance, photolyases are over-represented in the HRG set, highlighting photodamage as a major environmental pressure in the surface ocean.

There are, as might be expected, a number of categories with strongly positive and/or negative trends that will, as they are identified, be of help in identifying the nature of these HRGs, but many of them are of 'conserved unknown function'. Their identification will be essential in characterizing the microbes and the niches they inhabit.

Discussion

In this Article, we have referred to the ‘abundant’ and ‘cosmopolitan’ plankton as the ‘true’ picoplankton of the surface ocean. Here we distinguish between abundance in a sample and absolute numbers of cells. In many samples in the ocean, organisms with densities of

10^5 cells ml^{-1} or higher are abundant members of the community. Such levels are clearly not high in comparison with microbial densities seen in blooms, which may reach 10^7 cells ml^{-1} or higher: we observe 10^5 – 10^6 total microbes throughout the oceanic sites, and the ‘abundant’ microbes represent those taxa that are dominant in this otherwise organism-poor environment. This becomes an important part of the definition of the true picoplankton, as well-known microbial marine genera (for example *Vibrio*, *Alteromonas* and *Photobacterium*), which are surely present in low numbers, were rarely seen in our metagenomic studies. ‘Cosmopolitan’ implies that a given microbe was very often encountered at the surface sites studied, irrespective of the time of sampling. Because many of the microbes are not seen because of their low abundance, ‘abundant and cosmopolitan’ as judged by metagenomic data analysis is the descriptive term for what we call the ‘true’ picoplankton.

The rare and opportunistic biosphere

The work presented here reveals that most of the marine prokaryotic genomes that have been sequenced add little to our appreciation of the biology of the abundant and cosmopolitan picoplanktonic prokaryotes in the surface ocean, other than as a genomic contrast. We regard these low-recruiting microbes as members of a group that is adapted to niches other than the open ocean, including symbiotic, saprophytic, parasitic and other plant and animal associations. Members of this group are predicted to be strongly affected by eukaryotes and/or their products. They have the capacity to survive in the low-nutrient open ocean at low abundance levels (that is, residing in the 'long tail' of diverse organisms seen in the GOS data set), but are able to bloom if presented with the proper energy-rich conditions^{16,17}. Such organisms are known to use cell-cell communication mechanisms such as quorum sensing to regulate density-dependent processes such as biofilm formation and other 'group' activities. This notion predicts that the rapidly growing microbes, when found in their natural niches, may be clonal, growing from an original inoculation of one or a few cells. Such a prediction could be routinely tested by examining a number of energy-rich environments in the ocean.

In a recent paper¹⁸, it was proposed that the marine microbial world be divided into oligotrophs and copiotrophs, and that the former "dominate the ocean's free-living microbial populations". Our data support such a view, with the addition that the ability to regulate and adapt to changing conditions should allow the copiotrophs to survive long enough to find another nutrient-rich environment, be it a floating carcass, marine snow, a faecal pellet or the gut tract of a marine eukaryote. Indeed, if every eukaryotic species could harbour a number of species-specific bacterial associates, it would result in an immensely long tail of diversity: microbes awaiting their chance to bloom when conditions improved.

These thoughts are consistent with recent work¹⁹ in which an analysis of the deep-sea bacterial microbial metagenome (at 4,000-m depth near Hawaii) revealed a population more characteristic of the low-recruiting microbes reported here: that is, one rich in genes for motility, secondary metabolism, signal transduction, transport and other high-nutrient-type functions. A simple hypothesis, consistent with these data, is that the deep sea is simply too harsh for the monolithic surface picoplankton, and that the microbes here are again a reflection of a dynamic equilibrium between various high-nutrient environments. Given the sensory mechanisms and motility of the endogenous eukaryotes in the deep sea, they may have a major impact on the aquatic populations of bacteria and archaea observed there. Of particular note would be the issue of whether any of these bacteria and archaea might be of sufficiently low diversity that assembly of their genomes would be possible.

Genome modifications and streamlining

Our data suggest that genome streamlining of the cosmopolitan picoplankton is an important part of their success in the limiting environment of the surface ocean—an observation previously made²⁰ of members of the SAR11 clade. The few marine microbes that have been characterized suggest that the small genomes have retained virtually all of their biosynthetic abilities and a few key transport systems, but have dispensed with sensory/response systems (chemotaxis, quorum sensing and two-component regulators), motility, anaerobic metabolism and genes involved in organismal interaction. It seems that the marine picoplankton have a very low 'bacterial IQ'²¹. That is, there may be little need for rapid adaptation to changing conditions, with the result that expensive regulatory systems can be dispensed with. It is possible that these microbes, if they regulate their metabolism at all, may engage in regulation using specialized sigma factors, using non-coding RNAs or at the translational (or other post-transcriptional) level in preference to the more expensive transcriptional regulation characteristic of more rapidly growing and adaptive microbes.

This regulatory streamlining, combined with a metabolic single-mindedness, is consistent with the ability to survive as a free-living

organism in most of the surface ocean. It seems likely that the uncultured dominant and cosmopolitan picoplanktonic prokaryotes, such as SAR86, have similar overall genomic characteristics but with a different metabolic focus.

It is notable that the approach to genome streamlining seen here is fundamentally different from that seen in other very small microbes, as exemplified by recent work on the genome reduction in *Mycoplasma pneumonia*^{22–24}. Although that genome has been streamlined to a very small size, it has occurred without the loss of regulatory mechanisms needed for the bacterium's interactions with 'high-energy' and high-nutrient environments. Many closely related organisms in this group are obligate symbionts or pathogens, which are incapable of host-free growth. Another method of genome streamlining is seen in the recently reported widespread, nitrogen-fixing cyanobacterium UCYN-A²⁵. Like the *Mycoplasma* genomes, this small (~1.44-megabase) genome maintains many regulatory and metabolic abilities but discards many of its biosynthetic pathways, suggesting that it may be a symbiont with some as-yet-identified metabolic partner²⁶.

Cryptic escape

We put forward here a hypothesis that we call cryptic escape: success is achieved in the limiting oceanic environment by limiting the effective biomass in such a way as to discourage the success (and evolution) of specific predators, that is, by becoming 'invisible' to them as a food source. This can be done in two ways: by reducing population size and by reducing the amount of biomass per individual. With genome size well correlated with cell size (Supplementary Table 4), a population of SAR11 maintained at a density of 10^5 cells ml^{-1} has a biomass equivalent to less than 10^3 cells ml^{-1} *Vibrio* or *Shewanella*. Whether such a biomass could support the growth of a specific predator is not well documented, and would be an important test of the hypothesis. Cryptic escape thus views one of the driving forces of genome streamlining to be the avoidance of trophic predation, and would be consistent with many different functional (metabolic) end points, a prediction that will be directly tested as the genomes from other (currently uncultivated) cosmopolitan picoplankton groups become available.

It is equally important to know whether the maintenance of low cell numbers and slow growth rates would make viral predation inefficient enough that the CRISPR–Cas system is an unnecessary luxury for the abundant and cosmopolitan picoplankton. At a density of 10^5 cells ml^{-1} , each cell will be thousands of body lengths from others of its species and hundreds of body lengths from the nearest virus (not necessarily a virus specific to that cell), perhaps making efficient viral infection and growth a difficult prospect. With regard to viral predation, many other modes of defence might also be used, such as simply having many fewer metabolic receptors on their surfaces that viruses could use for recognition and entry, and/or extensive microdiversity within the dominant groups, such that annihilation by a single phage becomes nearly impossible. Each of these hypotheses is also testable with the right model systems.

Such an approach does not preclude the existence and success of general predators that can succeed by grazing non-specifically on picoplankters at densities as low as 10^5 – 10^6 cells ml^{-1} , which are often seen in oceanic waters. Thus, although the carbon fixed and cycled by the picoplankters is of great importance with regard to global carbon cycling, the low levels of biomass may have important consequences with regard to predation (both protistan and viral) of the very small plankters.

However, becoming small and/or rare is not within the purview of the picoplankters. When *Vibrio* (or other heterotrophs) are nutrient limited, they can become very small, often adopting a physiological state referred to as 'viable but not culturable'²⁷. Similar states can be seen for many normally fast-growing bacteria when they are nutrient limited at slow growth rates in a chemostat²⁸. Thus, it is well within the abilities of the larger heterotrophic bacteria to adopt the cryptic mode, and 'hide' within the long tail of diversity in the ocean.

Conclusion

In essence, the true single-celled picoplankton may be distinguished from most other prokaryotes more appropriately by what they cannot do than by what they can do. That is, the survival strategy seems to be one of dispensing with functions and/or with the control of functions. Thus, a picture emerges of microbes that survive by becoming small, single minded and uncommunicative. Furthermore, the streamlining of genomes is apparently a strategy that is available to many different taxonomic groups of microbes that make up the abundant and cosmopolitan picoplankton. Therefore, it is expected that the other uncultivated cosmopolitan picoplankton groups will show a similar form of genome streamlining, but with metabolic strategies distinct from the other HRGs.

METHODS SUMMARY

MMGSP genome sequencing, assembly and annotation. Two genomic libraries with respective insert sizes of 4 and 40 kilobases were made²⁹. The prepared plasmid and fosmid clones were sequenced from both ends on ABI 3730XL DNA sequencers (Applied Biosystems) at the JCVI Joint Technology Center to provide paired-end reads. Successful reads were assembled using the Celera Assembler³⁰ and the assembly was annotated using the JCVI prokaryotic annotation pipeline (<http://www.jcvi.org/cms/research/projects/annotation-service/overview/>).

Analysis of functional groups. The depth of coverage at $\geq 50\%$ nucleotide identity threshold (Supplementary Table 2) was used here. For each observed depth of coverage value, c , we binned the genomes into two groups (one containing genomes with coverage $< c$ and the other containing genomes with coverage $\geq c$), and assessed them using the Wilcoxon rank-sum test³¹ with the null hypothesis that the given protein category has the same distribution of values in the two groups. This assessment showed that, very frequently, the optimal (smallest) P value corresponded to a coverage of $c = 1$. On this basis, we grouped the 197 marine genomes into two sets: the HRG set, consisting of genomes with $c \geq 1$, and the LRG set, consisting of genomes with $c < 1$. We noted the P values for all protein categories (at $c = 1$), and computed a rejection threshold value (of 4.7×10^{-4}) after correction for multiple testing³² with the false-discovery rate set to 0.005. We repeated the P -value calculations for all protein categories after normalizing their original values (that is, raw counts) by dividing by the respective genome sizes (as given by the total number of proteins in the genome); here the rejection threshold value was 8.2×10^{-4} for the same false-discovery rate. Subsequently, of the 12,403 protein categories considered, 568 (4.6%) were identified as being differentially distributed in the HRG and LRG sets for both original and normalized data (Supplementary Table 3 and Supplementary Fig. 3), and were analysed further.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 18 December 2009; accepted 22 September 2010.

1. Pace, N. R. A molecular view of microbial diversity and the biosphere. *Science* **276**, 734–740 (1997).
2. Giovannoni, S. & Rappé, M. in *Microbial Ecology of the Oceans* (ed. Kirchman, D. L.) 47–84 (Wiley-Liss, Inc., 2000).
3. Rusch, D. B. *et al.* The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5**, e77 (2007).
4. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
5. Cole, J. R. *et al.* The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* **37**, D141–D145 (2009).
6. Yooshep, S., Li, W. & Sutton, G. Gene identification and protein classification in microbial metagenomic sequence data via incremental clustering. *BMC Bioinform.* **9**, 182 (2008).
7. Yooshep, S. *et al.* The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol.* **5**, e16 (2007).
8. Fischer, D. & Eisenberg, D. Finding families for genomic ORFans. *Bioinformatics* **15**, 759–762 (1999).
9. Scanlan, D. J. *et al.* Ecological genomics of marine picocyanobacteria. *Microbiol. Mol. Biol. Rev.* **73**, 249–299 (2009).
10. Bond, P. L., Hugenholtz, P., Keller, J. & Blackall, L. L. Bacterial community structures of phosphate-removing and non-phosphate-removing activated sludges from sequencing batch reactors. *Appl. Environ. Microbiol.* **61**, 1910–1916 (1995).
11. McCaig, A. E., Glover, L. A. & Prosser, J. I. Molecular analysis of bacterial community structure and diversity in unimproved and improved upland grass pastures. *Appl. Environ. Microbiol.* **65**, 1721–1730 (1999).
12. Rappé, M. S., Kemp, P. F. & Giovannoni, S. J. Phylogenetic diversity of marine coastal picoplankton 16S rRNA genes cloned from the continental shelf off Cape Hatteras, North Carolina. *Limnol. Oceanogr.* **42**, 811–826 (1997).

13. Dupont, C. L., Yang, S., Palenik, B. & Bourne, P. E. Modern proteomes contain putative imprints of ancient shifts in trace metal geochemistry. *Proc. Natl Acad. Sci. USA* **103**, 17822–17827 (2006).
14. Morel, F. M. M., Kustka, A. B. & Shaked, Y. The role of unchelated Fe in the iron nutrition of phytoplankton. *Limnol. Oceanogr.* **53**, 400–404 (2008).
15. Lis, H. & Shaked, Y. Probing the bioavailability of organically bound iron: a case study in the Synechococcus-rich waters of the Gulf of Aqaba. *Aquat. Microb. Ecol.* **56**, 241–253 (2009).
16. Nealson, K. H. & Hastings, J. W. Bacterial bioluminescence: its control and ecological significance. *Microbiol. Rev.* **43**, 496–518 (1979).
17. Nealson, K. H. & Venter, J. C. Metagenomics and the global ocean survey: what's in it for us, and why should we care? *ISME J.* **1**, 185–187 (2007).
18. Lauro, F. M. *et al.* The genomic basis of trophic strategy in marine bacteria. *Proc. Natl Acad. Sci. USA* **106**, 15527–15533 (2009).
19. Konstantinidis, K. T., Braff, J., Karl, D. M. & DeLong, E. F. Comparative metagenomic analysis of a microbial community residing at a depth of 4,000 meters at station ALOHA in the North Pacific subtropical gyre. *Appl. Environ. Microbiol.* **75**, 5345–5355 (2009).
20. Giovannoni, S. J. *et al.* Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**, 1242–1245 (2005).
21. Galperin, M. Y. A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. *BMC Microbiol.* **5**, 35 (2005).
22. Güell, M. *et al.* Transcriptome complexity in a genome-reduced bacterium. *Science* **326**, 1268–1271 (2009).
23. Kühner, S. *et al.* Proteome organization in a genome-reduced bacterium. *Science* **326**, 1235–1240 (2009).
24. Yus, E. *et al.* Impact of genome reduction on bacterial metabolism and its regulation. *Science* **326**, 1263–1268 (2009).
25. Zehr, J. P. *et al.* Globally distributed uncultivated oceanic N₂-fixing cyanobacteria lack oxygenic photosystem II. *Science* **322**, 1110–1112 (2008).
26. Tripp, H. J. *et al.* Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium. *Nature* **464**, 90–94 (2010).
27. Grimes, D. J., Mills, A. L. & Nealson, K. in *Nonculturable Microorganisms in the Environment* (eds Colwell, R. R. & Grimes, D. J.) 209–227 (ASM, 2000).
28. Ingraham, J. L., Maaloe, O. & Neidhardt, F. C. *Growth of the Bacterial Cell* 435 (Sinauer, 1983).
29. Goldberg, S. M. *et al.* A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. *Proc. Natl Acad. Sci. USA* **103**, 11240–11245 (2006).
30. Myers, E. W. *et al.* A whole-genome assembly of *Drosophila*. *Science* **287**, 2196–2204 (2000).
31. Mann, H. B. & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **18**, 50–60 (1947).
32. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to acknowledge the Gordon and Betty Moore Foundation (grant award agreement #521); the Department of Energy, Office of Science, Office of Biological and Environmental Research (DE-FC02-02ER63453); and the JCVI for funding this research. A.T. and S.Y. also acknowledge support from US National Science Foundation grant 0850256. We also thank the JCVI Software Engineering team for their contribution to the analyses, L. Sheehan for help with manuscript editing, and M. Thiagarajan, L. Foster, I. Singh and M. Lewis for help with data management. Finally, we thank the many collaborators who provided strains for sequencing, including R. Amann, F. Azam, D. Bartlett, R. Belas, C. Bolch, E. Bonch-Osmolovskaya, J. Bowman, I. Brettar, T. Brinkhoff, A. Buchan, R. Case, R. Cavicchioli, P. Chisholm, R. Edwards, D. Emerson, P. Falkowski, C. Francis, B. Fuchs, S. Giovannoni, J. Gonzalez, D. Green, A. Hagström, M. Hart, M. Haygood, R. Hill, M. Höfle, J.-C. Cho, K. Jürgens, S.-J. Kim, G. King, N. King, S. Kjelleberg, M. Klotz, M. Labrenz, E. Lilly, V. T. Marteinsson, M. A. Moran, A. Murray, J. C. Murrell, C. Pedros-Alio, J. Pinhassi, M. Polz, M. Rappé, A.-L. Reysenbach, F. Rodriguez-Valera, E. Rosenberg, N. Ruby, D. Scanlan, H. Schaefer, L. Shimkets, J. Siefert, M. Simon, T. Sokolova, E. Stabb, L. Stal, N. Tandeau de Marsac, B. Tebo, T. Thomas, C. Vetriani, I. Wagner-Dobler, B. Ward, J. Waterbury, B. Wawrik and A. Yayanos.

Author Contributions S.Y., K.H.N., D.B.R., J.P.M., S.J.W., A.T., G.S., A.E.A., C.L.D. and L.A.Z. analysed the data; S.F., K.B. and Y.H.R. sequenced the genomes; M.K., J.J. and R.M. assembled and annotated the genomes; S.Y., K.H.N., E.E., S.J.W., A.T., D.B.R., A.E.A., L.A.Z. and C.L.D. wrote the paper; and J.C.V., M.F. and R.F. were responsible for project management.

Author Information The sequences reported in this work have been deposited in the NCBI Entrez Genome Project database: individual accession numbers can be found in Supplementary Information. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike Licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.H.N. (knealson@usc.edu).

METHODS

MMGSP genome sequencing, assembly and annotation. Two genomic libraries with respective insert sizes of 4 and 40 kilobases were made²⁹. The prepared plasmid and fosmid clones were sequenced from both ends on ABI 3730XL DNA sequencers (Applied Biosystems) at the JCVI Joint Technology Center to provide paired-end reads. Successful reads were assembled using the Celera Assembler³⁰ and the assembly was annotated using the JCVI prokaryotic annotation pipeline (<http://www.jcvi.org/cms/research/projects/annotation-service/overview/>).

Fragment recruitment. BLAST³³ was used to recruit GOS reads to sequenced genomes³. For the purposes of computing the two recruitment statistics reported here, a recruited read was assigned to a single (best-matching) reference genome.

Alignment and tree building. Separate alignments of bacterial and archaeal sequences were produced using INFERNAL³⁴; bacterial and archaeal alignment models from the RDP⁵ were used for this purpose. Columns with gaps of >90% were removed from these alignments, and the two alignments were subsequently merged using MUSCLE³⁵. This alignment was used to construct a maximum-likelihood phylogeny using RAXML³⁶.

16S rRNA sequences from the RDP. We identified marine bacterioplankton 16S rRNA sequences from the RDP using a keyword search³⁷. Bacterial 16S rRNA sequences $\geq 1,200$ base pairs in length were downloaded and the text information in the GenBank records file was parsed. Only those records that contained word tokens from a positive control set and did not contain words from a negative control set were kept. The positive control set consisted of the following words: marine, coastal, ocean, sea, bacterioplankton, sar11, sar86, sar83, sar116, sar324, sar202. The negative control set consisted of the following words: deepsea, soil, sediment, sand, biofilm, freshwater, pond, lake, hydrothermal, groundwater, borehole, mud, petroleum, marinesnow, aquifer, halophil, oil, diesel, crust, anaerobe, symbiont, hygiene, rhizosphere, associated, viable, biofilter, reactor, sludge, gland, spleen, anoxic, spring, vent, volcanic, basalt, sponge, rock, bog, aquarium, benthic, bone, mat, marsh, mangrove, saltern, urchin.

16S rRNA sequences from GOS PCR libraries. Sequence data were generated using the protocol described in ref. 38. The list of GOS libraries is given in Supplementary Information. Chimaeric sequences were identified and removed. We used two programs to identify chimaeras: a modified version of the RDP chimaera checker⁵ and CHIMERASLAYER (<http://microbiomeutil.sourceforge.net>). Of the sequences that passed chimaera checking, only those $\geq 1,100$ base pairs in length were considered. These sequences were clustered using CD-HIT³⁹ at high identity and over nearly full length (that is, at $\geq 99\%$ identity over $\geq 95\%$ of the length of the shorter sequence). Only sequences in those clusters with five or more members were considered for further analysis; these comprised 37,860 sequences.

OTU identification. The RDP and GOS 16S PCR sequences were searched against 16S sequences from sequenced genomes using BLAST³³. Those sequences with an identity match of <97% were subsequently clustered using CD-HIT³⁹. The clustering was done successively at identities of 99%, 98% and 97%. Each 97%-identity cluster was considered to be an OTU, with size equal to the number of sequences in the flattened cluster. Each CD-HIT representative at 97% identity was considered to be the corresponding OTU's representative.

Determination of the phylogenetic distribution of the uncultured OTUs in 16S PCR samples. We clustered the unrecruited sequences using CD-HIT³⁹ at 97% identity to produce 1,493 OTUs, with the largest OTU containing 2,748 sequences and 93 OTUs containing ≥ 50 sequences. Sequences in these OTUs were classified to the class level using the RDP classifier⁵. Those OTUs that contained both RDP and GOS sequences, and those that contained GOS sequences from multiple libraries, were examined further. Representative sequences from these 320 OTUs were aligned, together with 16S sequences of the 197 marine genomes, using INFERNAL³⁴, and the alignment was used to construct a maximum-likelihood phylogeny using RAXML³⁶.

Searches against protein databases. The 739,579 proteins from the 197 marine genomes were searched for against several protein databases: COGs⁴⁰, Pfams⁴¹, TIGRFAMs⁴², KEGG pathways and modules⁴³, and MEROPS⁴⁴. For searches against COG profiles, a sequence was assigned to its best-matching COG; an *E* value cut-off of 10^{-8} was used. For Pfam and TIGRFAM assignments, only matches with scores above the model trusted cut-off score were considered. For KEGG assignments, sequences were assigned to KEGG pathways and modules on

the basis of best BLAST matches to genes in the KEGG orthologues collection. For MEROPS searches, sequences were assigned to the different families and clans using best BLAST matches to the peptidase units and inhibitor units included in MEROPS. For both sets of BLAST searches, only matches with bit scores of ≥ 50 were considered.

Analysis of functional groups. The depth of coverage at $\geq 50\%$ nucleotide identity threshold (Supplementary Table 2) was used here. For each observed depth of coverage value, *c*, we binned the genomes into two groups (one containing genomes with coverage $< c$ and the other containing genomes with coverage $\geq c$), and assessed using the Wilcoxon rank-sum test³¹ with the null hypothesis that the given protein category has the same distribution of values in the two groups. This assessment showed that, very frequently, the optimal (smallest) *P* value corresponded to a coverage of *c* = 1. On this basis, we grouped the 197 marine genomes into two sets: the HRG set, consisting of genomes with *c* ≥ 1 , and the LRG set, consisting of genomes with *c* < 1. We noted the *P* values for all protein categories (at *c* = 1), and computed a rejection threshold value (of 4.7×10^{-4}) after correction for multiple testing³² with the false-discovery rate set to 0.005. We repeated the *P*-value calculations for all protein categories after their original values (that is, raw counts) were normalized by dividing by the respective genome sizes (as given by the total number of proteins in the genome); here the rejection threshold value was 8.2×10^{-4} for the same false-discovery rate. Subsequently, of the 12,403 protein categories considered, 568 (4.6%) were identified as being differentially distributed in the HRG and LRG sets for both original and normalized data (Supplementary Table 3 and Supplementary Fig. 3), and were analysed further.

CRISPR–Cas systems. We searched for CRISPR arrays using a modified version of PILERC (http://www.drive5.com/pilercr/) and our own post-processing scripts for the removal of false-positives using MGTAXA (http://andreyto.github.com/mgtaxa). For the purposes of the current study, we computed several per-genome integral characteristics of the CRISPR system, such as the total number of Cas genes, the total number of CRISPR arrays and the minimal distance between any array and any Cas gene. The source code for the CRISPR analysis pipeline is available as part of our MGTAXA package. Independently of the metagenomic recruitment analysis, we compiled the known lifestyle information for two groups of microbial genomes. The first group consisted of all genomes possessing a significant CRISPR–Cas system (defined as at least three genes and three arrays per genome; *n* = 20), and the second consisted of 13 genomes randomly selected from the total set of genomes lacking CRISPR–Cas features (*n* = 130) (Supplementary Table 5). Within the first group, seven genomes were of deep-sea hydrothermal vent origin, two were from hypersaline environments and eight were associated with particles, surfaces or host organisms, or formed dense colonies in mats or blooms. By contrast, 12 out of 13 genomes lacking CRISPR–Cas were described as free-floating surface picoplankton.

33. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
34. Nawrocki, E. P., Kolbe, D. L. & Eddy, S. R. Infernal 1.0: inference of RNA alignments. *Bioinformatics* **25**, 1335–1337 (2009).
35. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
36. Stamatakis, A. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
37. Hagström, A. *et al.* Use of 16S ribosomal DNA for delineation of marine bacterioplankton species. *Appl. Environ. Microbiol.* **68**, 3628–3633 (2002).
38. Shaw, A. K. *et al.* It's all relative: ranking the diversity of aquatic bacterial communities. *Environ. Microbiol.* **10**, 2200–2210 (2008).
39. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
40. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**, 33–36 (2000).
41. Finn, R. D. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **36**, D281–D288 (2008).
42. Haft, D. H., Selengut, J. D. & White, O. The TIGRFAMs database of protein families. *Nucleic Acids Res.* **31**, 371–373 (2003).
43. Kanehisa, M. *et al.* KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36**, D480–D484 (2008).
44. Rawlings, N. D., Morton, F. R. & Barrett, A. J. MEROPS: the peptidase database. *Nucleic Acids Res.* **34**, D270–D272 (2006).

The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA

Josiane E. Garneau¹, Marie-Ève Dupuis¹, Manuela Villion¹, Dennis A. Romero², Rodolphe Barrangou², Patrick Boyaval³, Christophe Fremaux³, Philippe Horvath³, Alfonso H. Magadán¹ & Sylvain Moineau¹

Bacteria and Archaea have developed several defence strategies against foreign nucleic acids such as viral genomes and plasmids. Among them, clustered regularly interspaced short palindromic repeats (CRISPR) loci together with *cas* (CRISPR-associated) genes form the CRISPR/Cas immune system, which involves partially palindromic repeats separated by short stretches of DNA called spacers, acquired from extrachromosomal elements. It was recently demonstrated that these variable loci can incorporate spacers from infecting bacteriophages and then provide immunity against subsequent bacteriophage infections in a sequence-specific manner. Here we show that the *Streptococcus thermophilus* CRISPR1/Cas system can also naturally acquire spacers from a self-replicating plasmid containing an antibiotic-resistance gene, leading to plasmid loss. Acquired spacers that match antibiotic-resistance genes provide a novel means to naturally select bacteria that cannot uptake and disseminate such genes. We also provide *in vivo* evidence that the CRISPR1/Cas system specifically cleaves plasmid and bacteriophage double-stranded DNA within the proto-spacer, at specific sites. Our data show that the CRISPR/Cas immune system is remarkably adapted to cleave invading DNA rapidly and has the potential for exploitation to generate safer microbial strains.

Clustered regularly interspaced short palindromic repeats (CRISPR) loci were discovered in 1987 in *Escherichia coli*¹. However, the interest in these genetic elements increased in the early 2000s, as they were identified, along with many CRISPR-associated (Cas) proteins, in several prokaryotes^{2,3}. Recently, it was shown that the short spacers (21–72 base pairs (bp)) between these repeats originated from extrachromosomal DNA^{4–7}. Most importantly, it was experimentally demonstrated that those short spacers can provide resistance against bacteriophage infection and plasmid transformation^{8–10}.

The CRISPR/Cas immune systems act in at least two general steps: (1) the adaptation stage, where new spacers derived from foreign DNA (proto-spacers) are generally acquired at the leader end of the CRISPR locus^{11,12}; and (2) the interference stage, where the CRISPR/Cas system targets either invading DNA¹⁰ or RNA¹³. The mechanistic details of spacer acquisition are still unknown, but a clearer picture is emerging for the interference stage, which starts with the transcription of the CRISPR locus from a promoter located within the leader sequence^{14,15}. The full-length RNA is subsequently cleaved by a protein or protein complex, generating short CRISPR RNAs (crRNAs)^{16–19}. In *Pyrococcus*, Cas proteins use the crRNAs to target foreign RNA by complementarity in a ruler-anchored manner¹³. However, the *in vivo* mechanism of plasmid and viral interference has yet to be determined.

We previously showed that when bacteriophage-sensitive *Streptococcus thermophilus* cells are infected by virulent bacteriophages, a subset of cells (frequency of $<10^{-6}$) naturally diversify into bacteriophage-insensitive mutants through the acquisition of novel spacers derived from the invading bacteriophage genome into CRISPR1 and/or CRISPR3 (refs 9, 11, 20). Here we investigate the *in vivo* activity of the CRISPR/Cas system in *S. thermophilus* against both bacteriophage and plasmid DNA.

CRISPR/Cas affects plasmid stability

In silico analyses previously indicated that spacers may also be derived from *S. thermophilus* plasmids¹¹. Because the CRISPR/Cas system was recently demonstrated to interfere with plasmid transfer in *Staphylococcus*¹⁰, we put forward the hypothesis that the CRISPR/Cas system is responsible for the scarcity of plasmids in wild-type strains of *S. thermophilus*^{21,22}. To test this hypothesis, we first introduced by electroporation the vector pNT1 (ref. 23) into the plasmid-free *S. thermophilus* strain DGCC7710. This vector, which is derived from a native *S. thermophilus* plasmid, replicates via a rolling-circle mode and carries an added chloramphenicol-resistance (*cat*) gene as a selection marker. A representative transformant was grown in liquid medium for about 60 generations in the absence of chloramphenicol and aliquots were screened for antibiotic-sensitive colonies.

A total of 54 colonies (out of 900 tested) became sensitive to chloramphenicol and concomitantly lost pNT1. Thirty of these fifty-four colonies acquired a new spacer-repeat unit in CRISPR1. No spacer was integrated in the three other CRISPR loci of *S. thermophilus* DGCC7710 (ref. 8). Sequence analysis of CRISPR1 in the aforementioned 30 colonies identified 14 different spacers (S43–S56), all of which were homologous to pNT1 sequences (Table 1 and Fig. 1). Plasmid stability assays were also performed using two isogenic DGCC7710 strains in which *cas5* (*csn1*-like) or *cas7* genes associated to CRISPR1 were inactivated before the introduction of pNT1. Plasmid pNT1 was highly stable in the DGCC7710::*pcas5*[−] mutant as no chloramphenicol-sensitive colonies could be isolated, out of 1,800 screened. Of 170 randomly selected chloramphenicol-resistant colonies, none had acquired a new spacer in CRISPR1. On the other hand, chloramphenicol-sensitive colonies were readily obtained with the strain DGCC7710::*pcas7*[−], but none of the 200 colonies tested

¹Département de biochimie, de microbiologie et de bio-informatique, Faculté des sciences et de génie, Groupe de recherche en écologie buccale, Faculté de médecine dentaire, Félix d'Hérelle Reference Center for Bacterial Viruses, Université Laval, Québec City, Québec G1V 0A6, Canada. ²Danisco USA Inc., 3329 Agriculture Drive, Madison, Wisconsin 53716, USA. ³Danisco France SAS, Boîte Postale 10, F-86220 Dangé-Saint-Romain, France.

Table 1 | Plasmid-interfering mutants obtained during plasmid stability assay.

PIM	Spacer acquired* (5' to 3')	PAM†	Gene‡	Position on pNT1	Strand	Transformation with pNT1§ (CFU per µg DNA)	Number of times isolated
DGCC7710 _{pNT1} ^{+S43}	GATCAAATACTAATAAATACCCAGTACTT	TGCAGAAG	-	29	-	-	1
DGCC7710 _{pNT1} ^{+S44}	GACCCCCCTTTAAGTGCCGAGTGCCAAAT	TGAGAAA	<i>dso</i>	63	+	-	1
DGCC7710 _{pNT1} ^{+S45}	TATACTTGGGTAAATTATACCGTATGGCAA	AAGAAA	-	436	+	4.4 ± 3.5	1
DGCC7710 _{pNT1} ^{+S46}	TTTCCCAATCTTCTGGAATTGAATCGGGAT	AGAGTAG	<i>rep</i>	528	-	3.7 ± 2.6 × 10 ²	2
DGCC7710 _{pNT1} ^{+S47}	CATGATCTGCAATAATATTCAGACCTCGT	CTAGAAAT	<i>rep</i>	917	-	-	1
DGCC7710 _{pNT1} ^{+S48}	GATGATCTGCAATAATATTCAGACCTCGT	CTAGAAAT	<i>rep</i>	918	-	-	2
DGCC7710 _{pNT1} ^{+S49}	CGATGATCTGCAATAATATTCAGACCTCGT	CTAGAAAT	<i>rep</i>	919	-	-	1
DGCC7710 _{pNT1} ^{+S50}	AATTTAGTTCGTCAGTAGATTATGAAACT	GGAGAAG	<i>rep</i>	1074	+	-	1
DGCC7710 _{pNT1} ^{+S51}	AAAAGCAATGAGTTACATGGTTGCAAGAAT	GCAGAAA	<i>mob</i>	1488	+	-	4
DGCC7710 _{pNT1} ^{+S52}	GCCCCAGCTTACTATCAAGGAGCTTTCACG	GCATAAA	<i>sso</i>	1999	-	-	1
DGCC7710 _{pNT1} ^{+S53}	CGCCACAGGTTACTTGTCTCAAGGAGACC	ATCGAAT	-	2066	+	-	6
DGCC7710 _{pNT1} ^{+S54}	TCGTTTGTGTAACATAATGGGTGCTTTAGTT	GAAGAAT	-	2246	-	-	4
DGCC7710 _{pNT1} ^{+S55}	AGAGTTTTATGATTTATACCTTTCTGATGT	AGAGAAA	<i>cat</i>	2717	+	-	3
DGCC7710 _{pNT1} ^{+S56}	TTCTTCAACTAACGGGGCAGGTTAGTGACA	TTAGAAA	-	3114	-	-	2
DGCC7710	-	-	-	-	-	1.4 ± 0.6 × 10 ³	-

CFU, colony-forming units; PAM, proto-spacer adjacent motif; PIM, plasmid-interfering mutant.

* Nucleotide mismatch with pNT1 sequence is underlined.

† Nucleotide mismatches with the CRISPR1 proto-spacer adjacent motif (NNAGAAW) are underlined.

‡ The '-' symbol indicates an intergenic region.

§ The '-' symbol indicates that pNT1 could not be electroporated into the plasmid-interfering mutants (<1 CFU per µg DNA), $n \geq 2$.

|| The number of times that each plasmid-interfering mutant was isolated in the assay.

had acquired a new spacer in CRISPR1, indicating that plasmid loss was probably the result of other mechanisms responsible for plasmid instability^{24,25}. These data are consistent with previous findings that *cas5* is involved in the interference stage⁹ and its inactivation favours plasmid stability, whereas *cas7* is linked to the spacer acquisition stage⁹. Taken together, these observations indicate that the CRISPR/Cas system causes plasmid loss in *S. thermophilus*.

CRISPR/Cas targets antibiotic-resistance genes

To assess whether the acquired spacers from pNT1 cause plasmid interference in these plasmid-interfering mutants, we tested their propensity for pNT1 reintroduction. We were unable to obtain transformants of the plasmid-interfering mutant that carried a spacer perfectly matching a proto-spacer associated with the consensus proto-spacer adjacent motif NNAGAAW (Table 1). We previously showed that the newly added spacer must be identical to the proto-spacer in the bacteriophage genome to confer bacteriophage resistance, and that a proto-spacer adjacent motif located downstream of the proto-spacer is required for the resistance phenotype^{4,9,20,26}. Here, approximately half of the proto-spacer adjacent motif sequences contained one or two nucleotide mismatches. This tolerance for proto-spacer adjacent motif degeneracy could be due to the lower selective pressure for plasmids as compared to bacteriophages.

Some plasmid-interfering mutants targeting proto-spacers associated with non-consensus motifs (NNNAGAAG, NNATAAA, NNG GAAT or NNAGAAG) were also refractory to pNT1 reintroduction. Notably, a plasmid carrying the degenerate motif NNATAAA downstream of proto-spacer S52 could not be transformed into the corresponding plasmid-interfering mutants, whereas bacteriophages carrying the same proto-spacer adjacent motif could infect the matching bacteriophage-insensitive mutants²⁰ (Supplementary Table 1). It is also worth mentioning that one plasmid-interfering mutant contained a spacer (S47) that matched the last 29 nucleotides (out of 30) of the corresponding proto-spacer in pNT1, indicating that sequence identity at the 5' end of the spacer might be less important than in the middle or at the 3' end.

Two plasmid-interfering mutants (DGCC7710_{pNT1}^{+S55} and DGCC7710_{pNT1}^{+S56}, also named PIM S55 and PIM S56 for simplicity) carried a spacer targeting the *cat* gene (S55) or its downstream region (S56) (Table 1 and Fig. 1). These spacers rendered the cells untransformable with pNT1 and also with another plasmid (pTRK687) carrying the same *cat* gene. Thus, we conclude that the CRISPR/Cas system provides a simple and natural means to develop a bacterial strain that is refractory to the acquisition of plasmids that carry antibiotic-resistant genes.

Motif degeneracy influences plasmid interference

Two other plasmid-interfering mutants (PIM S45 and PIM S46) carrying a spacer associated with a non-consensus proto-spacer adjacent motif could be re-transformed with pNT1 but at lower frequencies (Table 1). Unexpectedly, the re-transformed PIM S46 visibly carried a linear form of pNT1 (Fig. 2a). *S. thermophilus* strain PIM S46 also contained the usual circular form of the vector, albeit at much lower concentration than the linear form (Fig. 2b), and could still grow in a medium containing chloramphenicol. However, 74–100% of the cells lost pNT1 within 14 generations after growth in a non-selective medium. Under the same conditions, 0–10% of the wild-type colonies lost the plasmid. Therefore, although the NNAGTAG motif was initially permissive for pNT1, the CRISPR/Cas machinery still eliminated the circular and linear plasmid forms within a few generations. The reason for the relative stability of the linear form of the plasmid, under selective pressure, is still unclear.

To assess whether the observed plasmid linearization was the result of CRISPR/Cas activity, the *cas5* and *cas7* genes of PIM S46 were also inactivated and the isogenic strains transformed with pNT1. Plasmid content analysis revealed the presence of only the circular form in PIM S46::*cas5*⁻, confirming the involvement of *cas5* in plasmid

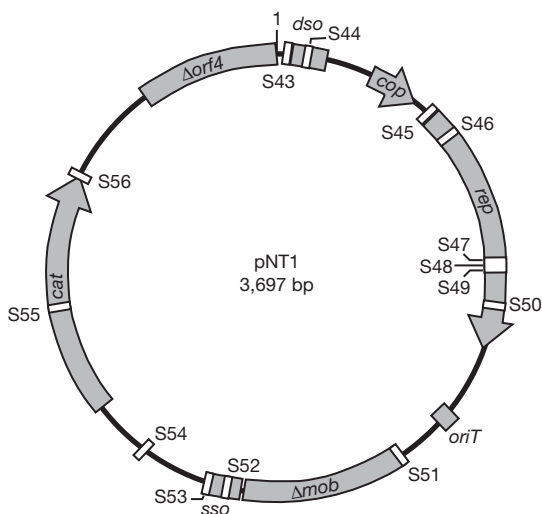


Figure 1 | pNT1 proto-spacers. The numbers outside and inside the map correspond to proto-spacers from the positive and negative strand, respectively.

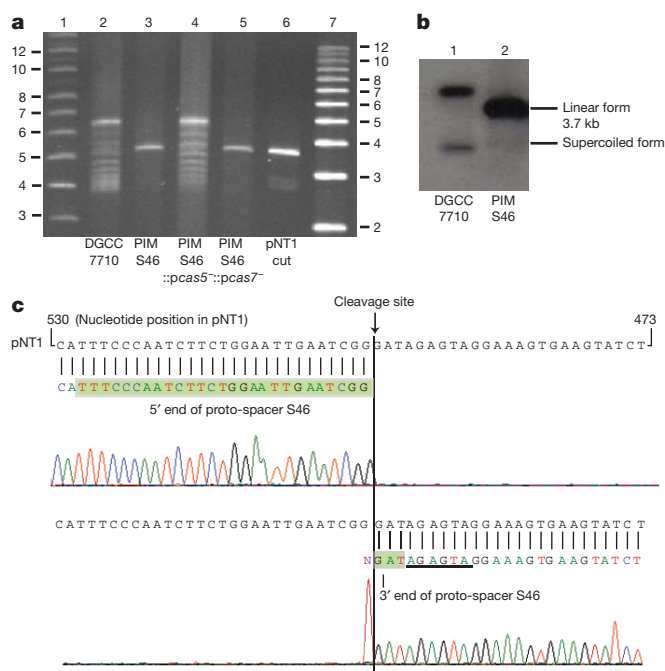


Figure 2 | The CRISPR/Cas system in *S. thermophilus* targets incoming plasmid DNA. **a**, Plasmid pNT1 in *S. thermophilus* strains. Twenty nanograms of plasmid DNA were loaded per lane. pNT1 plasmid was extracted from wild-type strain DGCC7710 (lane 2), from plasmid-interfering mutant (PIM) strain S46 (lane 3), PIM S46::pcas5⁻ (lane 4) and PIM S46::pcas7⁻ (lane 5). pNT1 from lane 2 was linearized with EcoRV (lane 6). Lanes 1 and 7: supercoiled and 1-kb DNA ladders (Invitrogen), respectively. **b**, Southern hybridization of plasmid pNT1 in *S. thermophilus* strains. Lane 1: 1 ng of native pNT1 plasmid extracted from wild-type strain DGCC7710. Lane 2: 10 ng of pNT1 extracted from PIM S46. The DIG-labelled probe 6 (Supplementary Table 2) was used. **c**, Direct sequencing electropherograms from primers NT17215_3 (upper part) and NT17225_2 (lower part) (see Methods). The non-templated addition of adenine (T in the reverse complement sequence shown here) at the extremity of the primer NT17225_2 sequence is a sequencing artefact due to the polymerase³⁵.

interference. On the other hand, pNT1 was still linear in mutant PIM S46::pcas7⁻ and in PIM S46 (Fig. 2a).

We next investigated the terminal ends of the pNT1 DNA molecules by directly sequencing the linear plasmid extracted from the PIM S46 strain. We unambiguously determined that pNT1 had been cut once (blunt), within the S46 proto-spacer, after the 27th nucleotide, 3 bases upstream of the proto-spacer adjacent motif (Table 2 and Fig. 2c). We conclude that double-stranded plasmid DNA is targeted and cleaved within the proto-spacer by the CRISPR/Cas machinery.

CRISPR/Cas system cuts viral DNA in the proto-spacers

To test whether the CRISPR/Cas system cleaves other invading DNA within the proto-spacer, we analysed the fate of bacteriophage DNA during the infection of bacteriophage-insensitive mutants containing

bacteriophage-derived spacers (Fig. 3a). We used the virulent streptococcal bacteriophage 2972 (ref. 27), its host DGCC7710, and three previously described bacteriophage-insensitive mutants derived from DGCC7710 following challenge with virulent bacteriophages 2972 and/or 858, namely DGCC7710₂₉₇₂^{+S4}, DGCC7710₂₉₇₂^{+S7} and DGCC7710₂₉₇₂^{+S4}₈₅₈^{+S32} (bacteriophage-insensitive mutant (BIM) S4, BIM S7 and BIM S4/S32, respectively)²⁰. These bacterial strains were infected with bacteriophage 2972 and total DNA was extracted from bacteriophage-infected cells sampled over time. Southern hybridizations of the DNA throughout the time course of infection of the bacteriophage-insensitive mutants with probes targeting regions close to the proto-spacers revealed that the bacteriophage DNA is rapidly cleaved by the CRISPR/Cas machinery (Fig. 3b, c). The cleavage site in the bacteriophage genome seemed to be within or in the vicinity of the proto-spacer during the infection of bacteriophage-insensitive mutants, whereas no cleavage of bacteriophage DNA was observed in the sensitive strain DGCC7710 (Fig. 3c). The cleavage patterns were similar in the bacteriophage-infected BIMs, even though the S4 and S7 spacers are derived from different DNA strands and transcriptional modules of the bacteriophage genome²⁸ (Fig. 3a). We also conducted the same experiments with isogenic derivatives of *S. thermophilus* BIM S4, in which the *cas5* and *cas7* genes were independently inactivated. The inactivation of *cas5* restored bacteriophage sensitivity (efficiency of plaquing of 1) and no bacteriophage DNA cleavage occurred, confirming the involvement of *cas5* in interference. In contrast, the inactivation of *cas7* did not affect bacteriophage resistance (efficiency of plaquing of 10⁻⁶) and bacteriophage DNA was cleaved (Fig. 3c).

We also investigated the fate of bacteriophage DNA during the infection of *S. thermophilus* BIM S4/S32, which contains two new spacers that target bacteriophage 2972. The bacteriophage genome was cleaved within each proto-spacer in this infected bacteriophage-insensitive mutant (Fig. 3c). The 5.2-kb and 5.8-kb bands corresponded to a cleavage at the S4 or S32 proto-spacers, respectively, whereas the 1.4-kb band resulted from cleavage at both sites. This is consistent with the previous observation that resistance increases with the number of spacers acquired²⁰. Our results show that the *in vivo* target of the *S. thermophilus* CRISPR/Cas system is DNA for both plasmid (Fig. 2) and bacteriophage (Fig. 3).

We then wanted to determine whether the bacteriophage DNA cleavage site was identical to that of the plasmid. Contrary to the method used for the determination of the cleavage site of plasmid pNT1, the low amount of cleaved bacteriophage DNA in the infected bacteriophage-insensitive mutants rendered direct sequencing of the extremities impossible. An alternative sequencing method was used (see Supplementary Methods and Supplementary Fig. 1), as the cleaved 1.4-kb fragment (from *S. thermophilus* BIM S4/S32) was subjected to inverse PCR to obtain a product that joins the ends of the S4- and S32-cleaved proto-spacers. Sequence analysis of the amplicon revealed that the bacteriophage fragment had indeed been cleaved within S4 and S32. For the other cleavage sites, the extracted bacteriophage DNA was digested *in vitro* by restriction endonucleases producing blunt ends

Table 2 | Cleavage sites within the proto-spacers in bacteriophage 2972 or plasmid pNT1

BIM or PIM	Spacer (size)	Homology (position)	Strand	Proto-spacer* (5' to 3')	PAM†
DGCC7710 ₂₉₇₂ ^{+S4}	S4 (30 nt)	Bacteriophage 2972 (31582)	+	CTCAGTCGTT ACTGGTGAACCAAGTTTC AAT	TGAGAAA
DGCC7710 ₂₉₇₂ ^{+S4} ₈₅₈ ^{+S32}	S4 (30 nt)	Bacteriophage 2972 (31582)	+	CTCAGTCGTT ACTGGTGAACCAAGTTTC AAT	TGAGAAA
DGCC7710 ₂₉₇₂ ^{+S4} ₈₅₈ ^{+S32}	S32 (30 nt)	Bacteriophage 2972 (33044)	+	ATTGTCTATTA CGACAACATGGAAGAT GAT	GTAGAAA
DGCC7710 ₂₉₇₂ ^{+S40}	S40 (29 nt)	Bacteriophage 2972 (31583)	+	TCAGTCGTT ACTGGTGAACCAAGTTTC AAT	TGAGAAA
DGCC7710 ₂₉₇₂ ^{+S7}	S7 (30 nt)	Bacteriophage 2972 (10299)	-	AAGCAAGTTGATATATTTCTCTTTCTT TAT	TAAGAAA
DGCC7710 ₂₉₇₂ ^{+S41}	S41 (30 nt)	Bacteriophage 2972 (31518)	-	TTCCCTTCGATAATGGCAAGACCGAAA CGT	TCAGAAA
DGCC7710 ₂₉₇₂ ^{+S42}	S42 (30 nt)	Bacteriophage 2972 (31084)	-	ATATTCATATTCCTGCTCATGTTTGA TAG	CAAGAAAT
DGCC7710 _{pNT1} ^{+S46}	S46 (30 nt)	Plasmid pNT1 (528)	-	TTTCCCAATCTTCTGGAATTGAATCGG GAT	AGAGTAG

BIM, bacteriophage-insensitive mutant; nt, nucleotide; PAM, proto-spacer adjacent motif; PIM, plasmid-interfering mutant.

* | Indicates the cleavage site

† The mismatches with the consensus proto-spacer adjacent motif (NNAGAAW) are underlined.

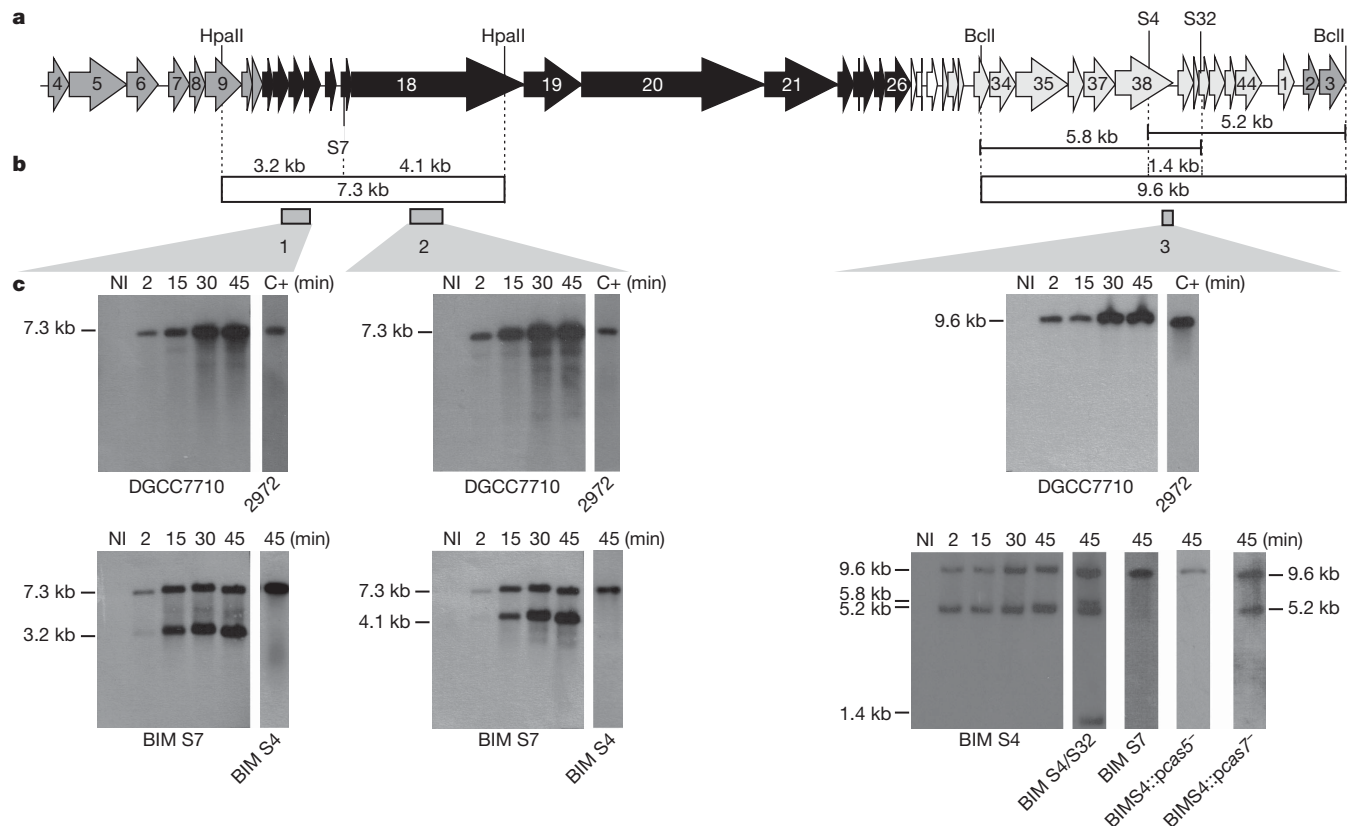


Figure 3 | The CRISPR1/Cas system targets bacteriophage DNA.

a, Bacteriophage 2972 genome and position of proto-spacers. Light grey, dark grey and black arrows indicate early, middle and late transcription module, respectively²⁸. Proto-spacers above and below the genome indicate positive and negative strands, respectively. **b**, Restriction fragments detected by hybridization and position of probes (1, 2 and 3). **c**, Southern blots of

and cutting upstream and downstream of each proto-spacer (see Supplementary Figs 1 and 2). Fragments comprising the cleaved proto-spacer at one extremity and a blunt site at the other were ligated and inverse PCRs were performed as described previously. PCR products were obtained for every fragment and sequenced.

Analysis of the amplicons revealed that all of the cleavage sites were within the proto-spacers (S4, S7 and S32). In every cleaved proto-spacer analysed, one cleavage site was located after the 27th nucleotide, 3 bases upstream of the proto-spacer adjacent motif, as observed for the linearized plasmid in PIM S46 (Table 2). Curiously, a second cleavage site was detected within the proto-spacers but only for the spacers targeting the positive strand of the bacteriophage genome (see S32 and S4 in Table 2). This second site was located 19 or 20 nucleotides upstream of the proto-spacer adjacent motif. Notably, the plasmid proto-spacer S46 described above was on the negative strand and was cut only once (Table 2). To confirm these results, three additional bacteriophage-infected BIMs that have acquired different spacers (S40, S41, or S42) were analysed similarly. The single (negative strand) and double (positive strand) cleavage sites within the proto-spacer were confirmed in these three bacteriophage-insensitive mutants (Table 2). Of note, the two cleavage sites within the S40 proto-spacer (a natural 5'-end truncated version of the S4 proto-spacer with 29 nucleotides instead of 30) were at the same position, indicating that the CRISPR1/Cas system in *S. thermophilus* probably acts in a 3'-end ruler-anchored manner. Ligation and amplification of each cleaved fragment could be obtained and no nucleotide was missing in the S7, S41, S42 and S46 proto-spacer sequences, confirming the blunt-end cleavage activity of *S. thermophilus* CRISPR1/Cas system.

Overall, we have established that the *S. thermophilus* CRISPR1/Cas system cleaves both bacteriophage and plasmid DNA *in vivo*. This

bacteriophage-infected strains DGCC7710, BIM S4, BIM S7, BIM S4/S32, BIM S4::pcas5⁻ and BIM S4::pcas7⁻ over 45 min (conducted at 2 min, 15 min, 30 min and 45 min). NI, non-infected strain. C+, positive control, 10 ng of digested DNA from bacteriophage 2972. One microgram of total DNA from DGCC7710 and 10 µg of bacteriophage-insensitive mutant per lane.

endonuclease activity, which seems to require *cas5*, is proto-spacer specific and orientation dependent. In that regard, the *S. thermophilus* CRISPR1/Cas interference mechanism differs from that recently described in *Pyrococcus furiosus*, whereby RNA is the target¹³ but corroborates the previous work on *Staphylococcus* which identified DNA as the target¹⁰. Furthermore, we show that the DNA cleavage activity is responsible for both bacteriophage resistance in bacteriophage-insensitive mutants and plasmid instability in plasmid-interfering mutants. The antagonism between plasmid maintenance and CRISPR/Cas activity probably explains the natural scarcity of plasmids in *S. thermophilus*. The CRISPR/Cas immune system can be used to naturally generate safer and more robust organisms with increased bacteriophage resistance and that are able to interfere with the dissemination of plasmids that carry antibiotic-resistance markers.

METHODS SUMMARY

Microbial conditions. *S. thermophilus* strains were grown at 37 °C or 42 °C in LM17 medium (ref. 20). Strains were infected with CsCl-purified bacteriophage 2972 at a multiplicity of infection of 5 (ref. 29). *cas5* and *cas7* genes were inactivated as described⁹. The identity of *S. thermophilus* strains was confirmed by sequencing the CRISPR loci⁸ and by bacteriophage assays²⁰.

Southern hybridization. DNA was extracted from bacterial cells as described³⁰. DNA probes for Southern blot hybridizations²⁹ were prepared with the PCR DIG labelling mix (Roche) (Supplementary Table 2). Hybridization and detection (CDP-star) were performed as recommended by Roche.

Plasmid stability. *S. thermophilus* DGCC7710 (or its *cas5*⁻ and *cas7*⁻ derivatives) transformed with pNT1 (ref. 23) was used to inoculate 10 ml of LM17. One-hundred microlitres of the previous culture was inoculated into 10 ml of fresh LM17 medium every morning (42 °C) and night (37 °C) for 5 days, for a total of 9 inoculations. For each culture, 100 colonies were screened for chloramphenicol sensitivity. The CRISPR1 of the chloramphenicol-sensitive clones was analysed.

Plasmid-interfering mutants with new spacers were electroporated³¹ with ~600 ng of pNT1 or pTRK687 (ref. 32). PIM S55 and PIM S56 were also transformed with pLS1 (ref. 33). Plasmid content of re-transformed PIM S45 and PIM S46 was verified by extraction (Qiagen) and extremities of the linearized plasmid from PIM S46 were sequenced.

Cleavage site of bacteriophage DNA. Total DNA from bacteriophage-infected bacteriophage-insensitive mutants was extracted³⁰ and digested using endonucleases generating blunt ends (Supplementary Fig. 2). After migration, agarose bands corresponding to both sides of the cleaved proto-spacers were recovered, DNA eluted and used for overnight intramolecular ligation at 16 °C, with T4 DNA ligase (Invitrogen). Ligation product was used as a template for inverse PCR³⁴ (Supplementary Figs 1 and 2 and Supplementary Table 3) and amplicons were sequenced (primers in Supplementary Table 3).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 April; accepted 20 September 2010.

- Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169**, 5429–5433 (1987).
- Jansen, R., van Embden, J. D. A., Gastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**, 1565–1575 (2002).
- Mojica, F. J., Díez-Villaseñor, C., Soria, E. & Juez, G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol. Microbiol.* **36**, 244–246 (2000).
- Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561 (2005).
- Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**, 174–182 (2005).
- Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**, 653–663 (2005).
- Andersson, A. F. & Banfield, J. F. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320**, 1047–1050 (2008).
- Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of Bacteria and Archaea. *Science* **327**, 167–170 (2010).
- Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
- Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843–1845 (2008).
- Horvath, P. et al. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1401–1412 (2008).
- Marraffini, L. A. & Sontheimer, E. J. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature Rev. Genet.* **11**, 181–190 (2010).
- Hale, C. R. et al. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**, 945–956 (2009).
- Lillestøl, R. K. et al. CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. *Mol. Microbiol.* **72**, 259–272 (2009).
- Pul, U. et al. Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. *Mol. Microbiol.* **75**, 1495–1512 (2010).
- Tang, T. H. et al. Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl Acad. Sci. USA* **99**, 7536–7541 (2002).
- Carte, J., Wang, R. Y., Li, H., Terns, R. M. & Terns, M. P. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **22**, 3489–3496 (2008).
- Hale, C., Kleppe, K., Terns, R. M. & Terns, M. P. Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* **14**, 2572–2579 (2008).
- Brouns, S. J. J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
- Deveau, H. et al. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400 (2008).
- Turgeon, N. & Moineau, S. Isolation and characterization of a *Streptococcus thermophilus* plasmid closely related to the pMV158 family. *Plasmid* **45**, 171–183 (2001).
- Somkuti, G. A. & Steinberg, D. H. Distribution and analysis of plasmids in *Streptococcus thermophilus*. *J. Ind. Microbiol.* **1**, 157–163 (1986).
- Vaillancourt, K. et al. Role of *galK* and *galM* in galactose metabolism by *Streptococcus thermophilus*. *Appl. Environ. Microbiol.* **74**, 1264–1267 (2008).
- Girard, S. L. & Moineau, S. Analysis of two theta-replicating plasmids of *Streptococcus thermophilus*. *Plasmid* **58**, 174–181 (2007).
- Kiewiet, R., Kok, J., Seegers, J. F., Venema, G. & Bron, S. The mode of replication is a major factor in segregational plasmid instability in *Lactococcus lactis*. *Appl. Environ. Microbiol.* **59**, 358–364 (1993).
- Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
- Lévesque, C. et al. Genomic organization and molecular analysis of virulent bacteriophage 2972 infecting an exopolysaccharide-producing *Streptococcus thermophilus* strain. *Appl. Environ. Microbiol.* **71**, 4057–4068 (2005).
- Duplessis, M., Russell, W. M., Romero, D. A. & Moineau, S. Global gene expression analysis of two *Streptococcus thermophilus* bacteriophages using DNA microarray. *Virology* **340**, 192–208 (2005).
- Sambrook, J. & Russell, D. W. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 2001).
- Fortier, L.-C. & Moineau, S. Morphological and genetic diversity of temperate phages in *Clostridium difficile*. *Appl. Environ. Microbiol.* **73**, 7358–7366 (2007).
- Buckley, N. D., Vadeboncoeur, C., LeBlanc, D. J., Lee, L. N. & Frenette, M. An effective strategy, applicable to *Streptococcus salivarius* and related bacteria, to enhance or confer electroporation competence. *Appl. Environ. Microbiol.* **65**, 3800–3804 (1999).
- Sturino, J. M. & Klaenhammer, T. R. Expression of antisense RNA targeted against *Streptococcus thermophilus* bacteriophages. *Appl. Environ. Microbiol.* **68**, 588–596 (2002).
- Nieto, C., Fernández de Palencia, P., López, P. & Espinosa, M. Construction of a tightly regulated plasmid vector for *Streptococcus pneumoniae*: controlled expression of the green fluorescent protein. *Plasmid* **43**, 205–213 (2000).
- Ochman, H., Gerber, A. S. & Hartl, D. L. Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**, 621–623 (1988).
- Clark, J. M. Novel non-templated nucleotide addition reactions catalyzed by prokaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res.* **16**, 9677–9686 (1988).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. R. Klaenhammer for pTRK687, B. Conway for editorial assistance, D. Tremblay, A. Fleury and B. Stahl for technical assistance. M.-E.D. is the recipient of a Novalait/FQRNT graduate scholarship and A.H.M. of a Clarin/FICYT postdoctoral scholarship. S.M. acknowledges funding from NSERC (Discovery program) and CIHR (Team Grant-Emerging: Novel Alternatives to Antibiotics).

Author Contributions S.M. conceived and headed the project. All the authors participated in the design of the study. J.E.G., M.-E.D., M.V. and A.H.M. performed the experiments. J.E.G. and S.M. wrote the main parts of the manuscript. All of the authors analysed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.M. (Sylvain.Moineau@bcm.ulaval.ca).

METHODS

Bacterial, bacteriophage and culture conditions. *Streptococcus thermophilus* strain DGCC7710 (ref. 9) and bacteriophage-insensitive mutants (BIMs) or plasmid-interfering mutants (PIMs) were grown in M17 broth supplemented with 0.5% lactose (LM17) at 37 °C or 42 °C. The identity of *S. thermophilus* strains was confirmed by sequencing the CRISPR loci⁸. Bacteriophage sensitivity of the isolated bacteriophage-insensitive mutants was estimated as reported previously³⁶. Bacteriophages were propagated in LM17 supplemented with 10 mM CaCl₂. Purified bacteriophage preparations (10¹¹–10¹² PFU ml⁻¹) were obtained by ultracentrifugation using a discontinuous CsCl gradient²⁹. *cas5* and *cas7* genes were inactivated as described previously⁹ in three *S. thermophilus* strains: DGCC7710, BIM S4 and PIM S46. All *cas5*⁻ and *cas7*⁻ derivatives were grown under erythromycin selective pressure. Efficiency of plaquing was determined as reported²⁰.

Bacteriophage infection. Bacterial strains were incubated at 42 °C in 100 ml of LM17 to an optical density of 0.5 at 600 nm. Cultures were concentrated by centrifugation and pellets were re-suspended in 10 ml of pre-incubated LM17 medium containing 10 mM CaCl₂. After removing a 1-ml uninfected sample, each bacterial culture was infected with purified bacteriophage 2972 at a multiplicity of infection (MOI) of 5, and incubated at 42 °C. Samples were taken after 2, 15, 30 and 45 min, centrifuged for 15 s at 16,000g and pellets were flash frozen, and stored at -80 °C until DNA extraction. The latent period of bacteriophage 2972 was previously determined to be 34 min on *S. thermophilus* DGCC7710. The maximum burst of bacteriophages is 40 min after the beginning of the infection²⁸. **DNA extraction and Southern hybridization.** Total DNA extractions were performed as described elsewhere³⁰, with these modifications: 25% (w/v) sucrose and 60 mg ml⁻¹ lysozyme were used as well as only one phenol-chloroform extraction. Dissolved DNA solutions were treated with 1 µg of RNase, incubated for 15 min at 37 °C, and their concentration determined using the Nanodrop 2000 spectrophotometer (Thermo Scientific).

Ten micrograms of bacteriophage-insensitive mutant DNA and 1 µg of wild-type DNA were digested with either HpaII or BclI (Roche). As controls, 10 µg of bacterial DNA and 10 ng of bacteriophage DNA were digested and were used for probe hybridizations. In addition, 500 ng of bacteriophage DNA was used as a digestion and transfer control. DNA fragments were separated in a 0.8% agarose gel in 1× TAE buffer, stained with EZ-Vision DNA dye (Amresco) and photographed under ultraviolet light. DNA was transferred onto positively charged nylon membranes (Roche) by capillary blotting²⁹. Membranes were stained for 45 s with methylene blue, rinsed with distilled water, and dried in ambient air. DNA probes were prepared with PCR DIG labelling mix (Roche) (Supplementary Table 2). Pre-hybridization, hybridization, washes and detection (CDP-star) were performed as recommended by Roche.

Plasmid stability and plasmid-interfering mutant analysis. *S. thermophilus* DGCC7710 (or the *cas5*⁻ and *cas7*⁻ isogenic derivatives) transformed with pNT1 plasmid²³ (GenBank accession number HQ010044) was picked from a LM17 plate containing 5 µg ml⁻¹ of chloramphenicol. Plasmid and CRISPR contents of this transformant were confirmed by sequencing and used to inoculate 10 ml of LM17. One-hundred microlitres of the previous culture was inoculated into 10 ml of fresh LM17 medium every morning and night for 5 days, for a total of 9 inoculations. Cultures were grown at 37 °C for overnight incubations and at 42 °C for 8 h during the day. Finally, these cultures were serially diluted and plated on LM17. For each culture, 100 colonies were screened for the absence of the antibiotic resistance phenotype. Plasmid stability of pNT1 was analysed by comparing the number of

colonies appearing on selective agar (5 µg ml⁻¹ of chloramphenicol) versus the number of colonies on LM17 agar. The CRISPR1 of chloramphenicol-sensitive clones was verified: the 5' end of the CRISPR1 of the plasmid-interfering mutants or bacteriophage-insensitive mutants was amplified by PCR with the primers yc70 (5'-TGCTGAGACAACCTAGTCTCTC-3')⁴ and RDS7revBamHI (5'-GGATCCGATCCGTTGAGGCCTTGTTTC-3'), and sequenced using the same primers. Fourteen different plasmid-interfering mutants having acquired a new spacer from pNT1 were electroporated as described elsewhere³¹ with ~600 ng of pNT1 or pTRK687 (ref. 32) vectors. All plasmid-interfering mutants (except for PIM S55 and PIM S56) could be transformed with the control vector pTRK687, which is carrying a chloramphenicol-resistance gene as selection marker. In the case of PIM S55 and PIM S56, they could be transformed with pLS1, which is carrying an erythromycin-resistance gene as a selection marker³³.

The proto-spacer region of pNT1 isolated from the pNT1-transformed PIM S45 was sequenced using the primers pNT1_104 (5'-GTGCCTTGAACCTTAGA GCCACAA-3') and NT17215_3 (5'-GTTCAGAGTATGGACTGCCG-3'). Chloramphenicol-resistant clones were also checked for spacer acquisition. Their CRISPR1 was amplified by PCR using the primers yc70 and RDS7revBamHI cited above.

Plasmid extraction and plasmid cleavage site determination. The plasmid content of *S. thermophilus* strains was isolated using QIAquick Spin Miniprep columns (Qiagen). The QIAquick protocol was modified by treating the cells with P1 buffer containing sucrose (25%) and lysozyme (60 mg ml⁻¹), and incubating at 37 °C for 15 to 30 min. Extremities of the linearized plasmid were directly sequenced with the primers NT17225_2 (5'-TACGTACTTGTGTTACTA TTG-3') and NT17215_3 (5'-CGGCAGTCCATACTCTGAAC-3').

Bacteriophage cleavage-site determination. Each bacteriophage-insensitive mutant was independently infected for 45 min with bacteriophage 2972 at a multiplicity of infection of 5. Total DNA from bacteriophage-infected bacteriophage-insensitive mutant was extracted³⁰ and digested using endonucleases cutting upstream and downstream of each proto-spacer and generating blunt ends (Roche and NEB) (Supplementary Fig. 2). DNA was migrated on a 0.8% agarose gel. Bands corresponding to the left and right fragments of the cleaved proto-spacer were extracted from the gel (band sizes estimated according to cleavage within the proto-spacer). DNA was extracted using the QIAEX II Gel Extraction Kit (Qiagen). Each fragment comprised the cleaved proto-spacer at one extremity and a blunt site at the other. Total eluted product (50 µl) of each fragment was used for overnight intramolecular ligation at 16 °C, with T4 DNA ligase (Invitrogen). In the case of BIM S4/S32, the fragment delimited by the cleaved proto-spacers S4 and S32 could be directly used for intramolecular ligation. Five microlitres of the ligation product were used as a template for PCR (100 µl reaction volume), with Taq Polymerase (Roche) (Supplementary Table 3). The cycling conditions were: 94 °C/45 s, 55 °C/45 s, 72 °C/1 min for 35 cycles with a final extension of 72 °C/10 min³⁴. PCR products were sequenced using primers listed in Supplementary Table 3.

Sequences. All sequencing reactions were performed by the 'Plateforme de Séquençage et de Génotypage des Génomes' service from the CHUL-CHUQ Research Center. The sequences from both strands were analysed using BioEdit software (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>) or Staden package (http://staden.sourceforge.net/staden_home.html).

36. Moineau, S. et al. Characterization of lactococcal bacteriophages from Québec cheese plants. *Can. J. Microbiol.* **38**, 875–882 (1992).

Video imaging of walking myosin V by high-speed atomic force microscopy

Noriyuki Kodera^{1,2}, Daisuke Yamamoto^{1,2}, Ryoki Ishikawa³ & Toshio Ando^{1,2}

The dynamic behaviour of myosin V molecules translocating along actin filaments has been mainly studied by optical microscopy. The processive hand-over-hand movement coupled with hydrolysis of adenosine triphosphate was thereby demonstrated. However, the protein molecules themselves are invisible in the observations and have therefore been visualized by electron microscopy in the stationary states. The concomitant assessment of structure and dynamics has been unfeasible, a situation prevailing throughout biological research. Here we directly visualize myosin V molecules walking along actin tracks, using high-speed atomic force microscopy. The high-resolution movies not only provide corroborative 'visual evidence' for previously speculated or demonstrated molecular behaviours, including lever-arm swing, but also reveal more detailed behaviours of the molecules, leading to a comprehensive understanding of the motor mechanism. Our direct and dynamic high-resolution visualization is a powerful new approach to studying the structure and dynamics of biomolecules in action.

Proteins are dynamic in nature and work at the single-molecule level. Reflecting this fact, single-molecule fluorescence microscopy has been widely exploited to understand how proteins operate (reviewed in ref. 1). However, what we can observe thereby is the dynamic behaviour of individual fluorescent spots (each being emitted from a fluorophore attached to a selected locus of the molecule), not of the protein molecules themselves. The structure of proteins has been studied by electron microscopy, X-ray crystallography and NMR, but the obtained structures are essentially static. To overcome this long-standing problem and make it possible to record the structure and dynamics of functioning biomolecules simultaneously, high-speed atomic force microscopy (HS-AFM) has been developed^{2–5}. The recent significant improvement in its performance has been demonstrated in a few imaging studies of proteins^{6–8}.

Myosin V is a two-headed processive motor and functions as a cargo transporter in cells (reviewed in ref. 9). Facilitated by the processivity of the molecule^{10,11}, numerous single-molecule studies have demonstrated that it moves hand over hand with a ~ 36 -nm advance^{12,13} for every adenosine triphosphate (ATP) hydrolysis¹⁴. However, a comprehensive description of the molecule in action has not yet been attained. We applied HS-AFM to this highly dynamic protein to capture its minute dynamic behaviours. In addition to known and speculated behaviours, previously unappreciated ones clearly appeared in the molecular movies: 'stomping-like' brief detachment and rebinding of either head; occasional unwinding of the coiled-coil tail of the two-headed bound molecule followed by the leading lever-arm swing, indicating the presence of intramolecular tension responsible for the power stroke; and repeated conformational switching of the leading head by adenosine diphosphate (ADP) binding and release. These findings lead to a better understanding of how myosin V operates for its motor function.

Unidirectional processive movement

Partially biotinylated actin filaments were immobilized using streptavidin with a low surface density on biotin-containing lipid bilayers formed on a mica surface (Supplementary Fig. 1). To facilitate the

weak sideways adsorption of tail-truncated myosin V (M5-HMM) onto the bilayer surface, a positively charged lipid (1,2-dipalmitoyl-3-trimethylammonium-propane (DPTAP), 5%) was included in the bilayer unless otherwise stated. This condition was necessary to clearly visualize M5-HMM molecules.

Figure 1a shows images of M5-HMM moving processively with discrete ~ 36 -nm steps (Supplementary Movie 1). A processive run almost always continued to the full extent of each imaging range used (that is, 4–16 steps). Typical long runs tracked by shifting the scan area are shown in Supplementary Movie 1c, d. In the two-headed bound M5-HMM (where by head we mean the motor domain and the neck domain), the neck-motor domain junction appears smooth in the leading head (L-head) but is V-shaped in the trailing head (T-head) without exception (Fig. 1b), because the neck regions emerge from different parts of the motor domain. The short coiled-coil tail was mostly tilted towards the minus end of actin (Fig. 1a, arrows). These features are consistent with electron microscopy observations^{15–17} and can be used to determine the actin polarity when bound M5-HMM is stationary.

The average translocation velocity (V) as a function of ATP concentration ($[ATP]$) was well fitted by $V = d_s(1/k_1[ATP] + 1/k_2)^{-1}$, where d_s is the step size, k_1 is the second-order ATP binding rate constant and k_2 is the first-order ADP dissociation rate constant (Supplementary Fig. 2). The maximum velocity observed without DPTAP was similar to that measured by fluorescence microscopy^{13,18}, indicating the negligible effects of the tip-sample and surface-sample interactions on motor activity. When 5% DPTAP was included in the substrate surface, the values $k_1 = 0.9 \pm 0.3 \mu M^{-1} s^{-1}$ and $k_2 = 7.2 \pm 1.5 s^{-1}$ were obtained (Supplementary Table 1). The former agrees with previous reports, but the latter is about 70% of previously reported values^{13,18} ($\sim 11 s^{-1}$), suggesting that the surface-sample interaction slightly reduces the rate of ADP release from the T-head.

Hand-over-hand movement

As observed above, the ~ 36 -nm advance was completed within one frame (146.7 ms); therefore, the molecular process occurring during a

¹Department of Physics, Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan. ²CREST, JST, Sanban-cho, Chiyoda-ku, Tokyo 102-0075, Japan. ³Department of Molecular and Cellular Pharmacology, Gunma University Graduate School of Medicine, 3-39-22 Showa-machi, Maebashi 371-8511, Japan.

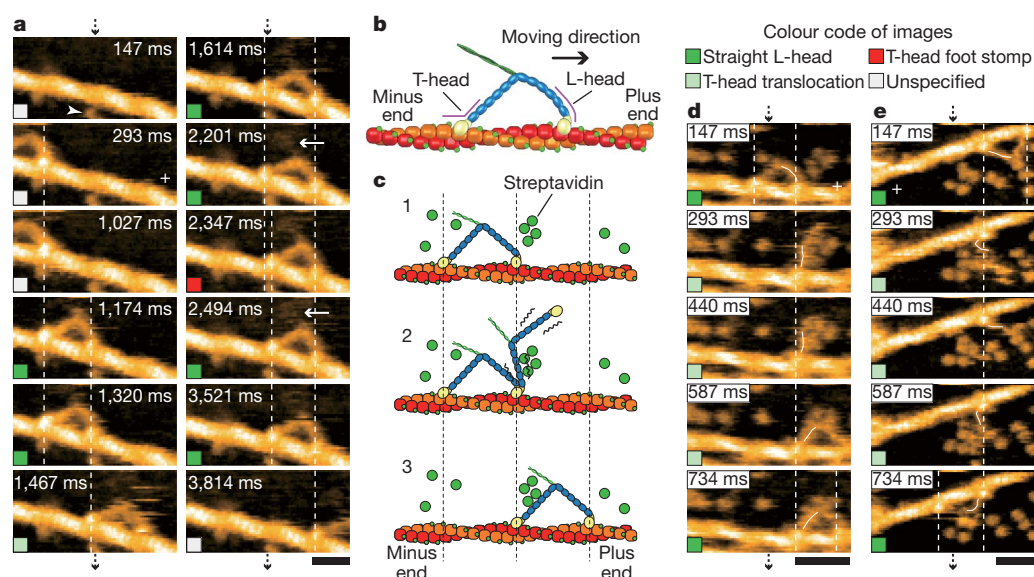


Figure 1 | Directly visualized walking M5-HMM. **a**, Successive AFM images showing processive movement of M5-HMM in 1 μ M ATP (Supplementary Movie 1). Arrowhead, streptavidin molecule; arrows, coiled-coil tail of M5-HMM tilted towards the minus end of actin. Scan area, $130 \times 65 \text{ nm}^2$; scale bar, 30 nm. **b**, Schematic of two-headed bound M5-HMM. **c**, Schematic explaining the images in **d** and **e**. **d**, **e**, Successive AFM images showing hand-over-hand

movement in 1 μ M ATP (**d**; scan area, $150 \times 75 \text{ nm}^2$; scale bar, 50 nm) and in 2 μ M ATP (**e**; scan area, $130 \times 65 \text{ nm}^2$; scale bar, 30 nm) (Supplementary Movie 2). The swinging lever is highlighted with a thin white line. Vertical dashed lines in **a**, **d** and **e** show the centres of mass of the motor domains, and a plus sign is used to indicate the plus end of actin. All images were taken at a 146.7 ms per frame.

step could not be resolved. However, additional streptavidin molecules placed on the substrate surface as moderate obstacles to the advance allowed the visualization of this process (Fig. 1c–e). After T-head detachment, the nearly straight leading neck swung from the reverse arrowhead (R-ARH) orientation to the arrowhead (ARH) orientation (Supplementary Movie 2), confirming the swinging lever-arm motion initially proposed for muscle myosin¹⁹. The detached T-head rotationally diffused around the advancing neck-neck junction (no translational diffusion on the actin occurs; Supplementary Movie 2c) and then bound to a forward site on the actin filament, completing one step. Thus, the hand-over-hand movement^{12,13}, including the intermediate process, was directly visualized at high resolution. We did not observe other behaviours of M5-HMM predicted by the ‘inchworm’²⁰ and ‘biased diffusion’²¹ models, at least with the time resolution used. The captured images show that the forward movement is driven not by bending but by rotation of the L-head. The rotation seems to occur spontaneously after T-head detachment, suggesting that intramolecular tension driving the L-head swing exists in the two-headed bound molecules.

Unfolding of the coiled-coil tail

In a saturating amount of ADP, both heads were bound to a single actin filament and the L-head always appeared nearly straight (curved slightly outwards; Fig. 2a, Supplementary Movie 3 and Supplementary Fig. 3b), as observed for M5-HMM moving in 0.1 μ M and 1 μ M ATP. In this nearly straight conformation, the neck domain appears to emerge from the rear of the motor domain¹⁷. During imaging of the two-headed bound M5-HMM, we sometimes encountered unwinding of the short coiled-coil tail. After unwinding, the monomerized L-head immediately rotated towards the ARH orientation (Fig. 2b and Supplementary Movie 4). This head separation was rarely reversed (Supplementary Movie 4c, d). This unwinding is unlikely to be caused by the tip-sample interaction because monomeric heads were detected in the earliest frames in the successive movies in new scan areas (see the latter half of Supplementary Movie 4b), and because a previous study showed the presence of a single-head fragment with an intact coiled-coil region in an expressed M5-HMM sample, although it was interpreted to be due to partial proteolysis²². Rather, this finding directly demonstrates that intramolecular tension exists in the

two-headed bound M5-HMM, and that the tension release caused by the unwinding results in the L-head rotation, similar to the L-head behaviour in the hand-over-hand movement. These observations provide the following insights into the motor mechanism: first, the neck-motor domain junction of L-head-ADP is sufficiently flexible to allow the head to bind strongly to actin in the R-ARH orientation, even without transitioning through a weak-binding ADP- P_i -bound state (P_i , inorganic phosphate); second, this strong binding forces the L-head into a strained ‘pre-stroke’ state; and third, the rotation of the L-head is likely to occur spontaneously when the constraint holding it in the strained state is removed by T-head detachment, suggesting that the lever swing by the L-head is not accompanied by chemical transitions there.

Foot stomp in ATP

In general, during the actomyosin ATPase cycle, the strained pre-stroke state is always formed after P_i is released from an ADP- P_i -bound head weakly interacting with actin²³. Thus, it may be considered that the strained pre-stroke state, which is formed directly by the binding of L-head-ADP to actin, does not participate in the forward step. However, we observed molecular behaviour in ATP that indicates that it does participate. In two-headed bound M5-HMM, both of the motor domains frequently exhibited brief dissociation and reassociation on the same actin filament (or a brief translocation by around $\pm 5 \text{ nm}$ along the actin filament), whereas M5-HMM remained at approximately the same position on the filament (Fig. 1, Supplementary Fig. 3a, Supplementary Movies 1 and 2, and Methods). We have termed this behaviour ‘foot stomp’. Note that when the T-head dissociates, it rarely rebinds to actin without stepping forward; thus, the T-head foot stomp is usually observed as a brief translocation along the actin filament. The foot stomp was more frequently observed at the L-head than at the T-head (by a ratio of approximately 3:1; Fig. 3a). The briefly detached L-head does not carry bound P_i because P_i is rapidly released from an ADP- P_i -bound head when it attaches to actin²⁴. Nevertheless, the detached L-head with only ADP bound rebinds to actin, still in the R-ARH orientation, and then swings forward following T-head detachment.

At each step, the L-head foot stomp occurred about twice in 0.1 μ M ATP and 0.4 times in 1 μ M ATP on average. These high frequencies of

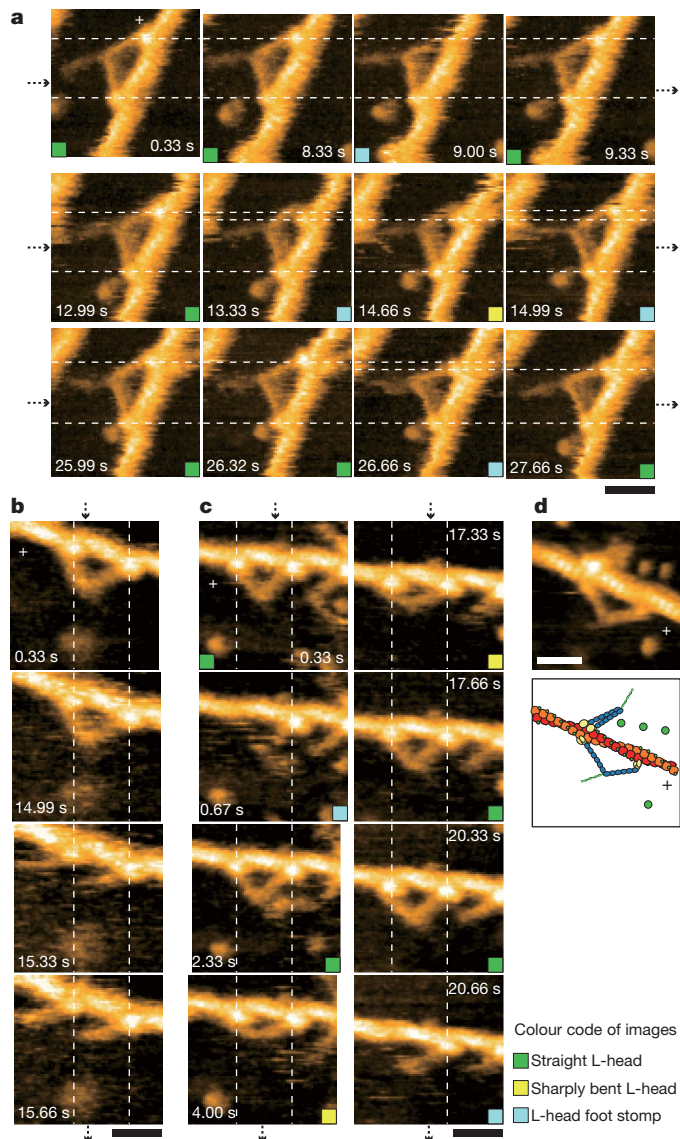


Figure 2 | AFM images of M5-HMM in ADP and under the nucleotide-free condition. **a**, Successive AFM images taken in 50 μM ADP (Supplementary Movie 3). **b**, Successive AFM images showing unwinding of the short coiled-coil tail of two-headed bound M5-HMM observed in 50 μM ADP (Supplementary Movie 4). **c**, Successive AFM images made under the nucleotide-free condition (Supplementary Movie 5). **d**, AFM image and illustration of nucleotide-free M5-HMM with heads bound to adjacent actin subunits. A plus sign is used to indicate the plus end of actin. All images were taken at 333.2 ms per frame with a scan size of $90 \times 90 \text{ nm}^2$. Scale bars, 30 nm.

occurrence indicate that a foot stomp at the L-head occurs while it is bound to ADP, not while it is bound to ADP- P_i . If P_i release, the rate of which is independent of [ATP], were significantly delayed by the interaction of M5-HMM with the substrate surface containing 5% DPTAP, M5-HMM in high [ATP] would walk much more slowly than observed. Moreover, the foot stomp also occurs when DPTAP is absent from the substrate. A foot-stomp-like behaviour was previously suggested in a fluorescence microscopy observation of walking myosin V molecules²⁵. Thus, the foot stomp is an inherent behaviour of myosin V. As will be described later, the foot stomp also occurs in ADP and under the nucleotide-free condition.

Conformational transition in the L-head

Even under the nucleotide-free condition, M5-HMM was bound to actin through the two heads (Fig. 2c, Supplementary Movie 5 and Supplementary Fig. 3c). However, unlike in the presence of nucleotides,

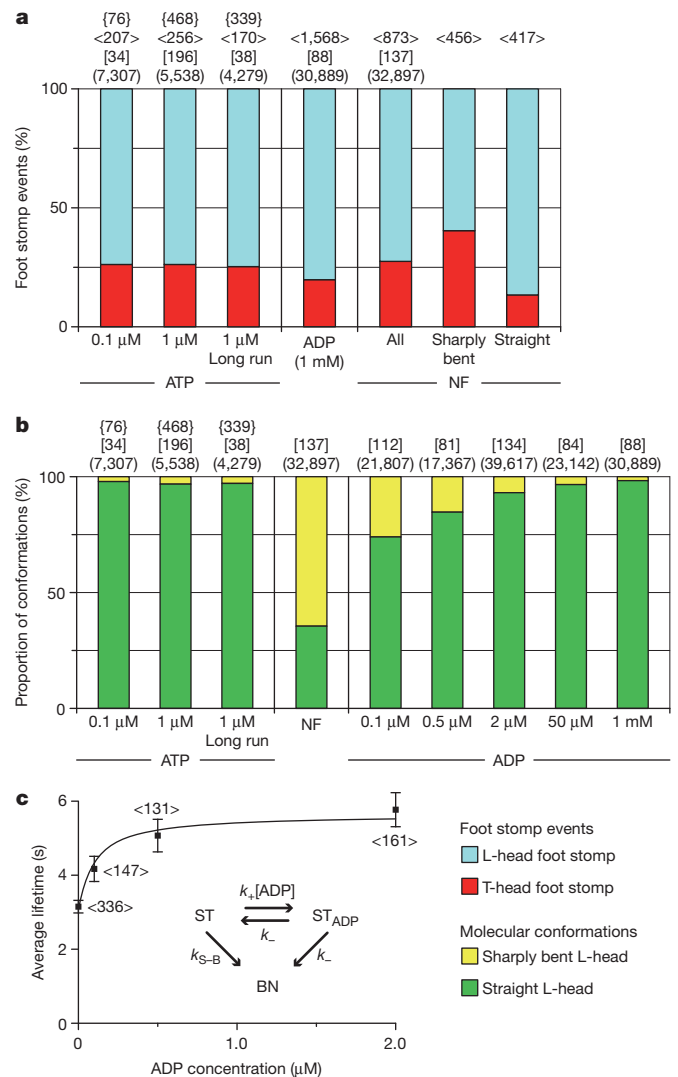


Figure 3 | Foot stomp and L-head conformations in different nucleotide conditions. **a**, **b**, Proportion of foot stomp frequencies observed at the L- and T-heads (**a**) and of the detected straight and sharply bent conformations of the L-head (**b**). The numbers of steps, events, molecules and frames examined using images within which both heads were fully located are enclosed in curly brackets, $\langle \text{angle} \rangle$, [square] and (round) brackets, respectively. 'Long run' represents data obtained using wide imaging areas. NF, nucleotide free. **c**, Average lifetimes of the L-head with a straight conformation for various [ADP] values. The total numbers of straight-to-bent transition events observed with ~ 50 different molecules under the respective [ADP] values are shown in angle brackets. Inset, reaction scheme (ST, nucleotide-free, straight L-head; ST_{ADP} , ADP-bound, straight L-head; BN, sharply bent, nucleotide-free L-head). Error bars, $\pm \text{s.e.m.}$; $k_{\text{S-B}} = 0.32 \pm 0.02 \text{ s}^{-1}$ (best-fit value $\pm \text{s.d.}$).

the leading neck frequently had a sharply bent conformation and alternated back and forth between this conformation and the nearly straight conformation (Fig. 2c and Supplementary Fig. 3c), indicating that they are in equilibrium. The sharply bent conformation at the L-head was also observed recently by negative-stain electron microscopy of actin-M5-HMM, although the chemical state was unspecified¹⁷. In the sharply bent conformation, both heads have the ARH orientation around the motor domain; thus, both necks appear to emerge from the front of the respective motor domains. As will be clarified later, this sharp bending has no relevance to power-stroke generation but provides a useful indicator of whether or not the L-head contains ADP.

The proportions of the straight L-head detected (r) were 0.98 in 1 mM ADP and 0.36 under the nucleotide-free condition (Fig. 3b). This finding is consistent with the ARH orientation distributions of single-headed

myosin V under the nucleotide-free and ADP-bound conditions (Supplementary Fig. 4). Under the nucleotide-free condition, the orientation angle relative to the long axis of the actin filament was $34 \pm 10^\circ$ (mean \pm s.d.) (although a double-Gaussian distribution is somewhat significant: $33 \pm 9^\circ$ and $51 \pm 7^\circ$; F -test, $P < 0.05$). In ADP, however, the angle was distributed more widely ($29 \pm 11^\circ$ and $51 \pm 15^\circ$; F -test, $P < 10^{-6}$), which may be relevant to two different ADP-bound states in equilibrium²³. Therefore, the hinge around the neck-motor domain junction of the ADP-bound head is more flexible than that of the nucleotide-free head. Because of the relatively rigid hinge of the nucleotide-free head, the leading neck in the R-ARH orientation tends to be sharply bent to release the large strain accumulated in the neck. This rigid hinge is also indicated by the observation that both heads are rarely bound to adjacent actin subunits predominantly in the nucleotide-free condition (Fig. 2d).

ADP release from the L-head

In 1 μM ATP and even in 0.1 μM ATP, the leading neck was mostly straight ($r \approx 0.98$) (Fig. 3b), which suggests that the L-head retains ADP until the T-head binds to ATP and detaches from actin, consistent with previous reports^{14,26–30}. From the proportion (Fig. 3b) and lifetime (Fig. 3c and Supplementary Movie 6) of the straight L-head as a function of [ADP], we estimated the kinetic parameters for ADP binding/dissociation on the L-head: the ADP dissociation constant is $K_d = k_-/k_+ = 0.075 \pm 0.013 \mu\text{M}$; the ADP dissociation rate constant is $k_- = 0.100 \pm 0.004 \text{ s}^{-1}$; and the second-order ADP binding rate constant is $k_+ = 1.3 \pm 0.3 \mu\text{M}^{-1}\text{s}^{-1}$. The value of k_- is ~ 70 times smaller than the corresponding rate for the T-head estimated above ($k_2 = 7.2 \pm 1.5 \text{ s}^{-1}$). This degree of asymmetry between the two heads is similar to that previously reported²⁶. The value of k_+ is ~ 10 times smaller than the corresponding value, $12.6 \mu\text{M}^{-1}\text{s}^{-1}$, previously measured in a solution study on single-headed myosin V²⁴. The ADP dissociation rate constant of 0.100 s^{-1} means that, on average, one ADP is released from the L-head every 10 s. However, M5-HMM walks many steps during 10 s. Thus, we confirm the consensus view that the sequential events of ADP release, the subsequent ATP binding and the resulting head dissociation take place solely at the T-head, which is the basis for the processivity and hand-over-hand stroke generation^{14,26–30}. Accordingly, the sharp bending observed at the nucleotide-free L-head has no relevance to power-stroke generation.

Foot stomp in ADP and nucleotide-free conditions

The foot stomp observed in ATP also occurred with higher frequencies at the L-head under all the nucleotide conditions examined here (Fig. 3a, Supplementary Fig. 5 and Supplementary Movies 3 and 5). Note that in ADP and under the nucleotide-free condition, complete detachment of the T-head was only sometimes observed. This detachment was followed by a forward step. This event was also counted as a foot stomp. The foot stomp at the T-head mostly occurred as brief translocation along actin. The frequency difference between the two heads became smaller when the L-head assumed a sharply bent conformation (L-head/T-head frequency ratio, $\sim 3:2$). Thus, the frequency of foot stomping strongly correlates with the orientation of the bound head. Both heads would by nature preferentially take the ARH orientation. However, the nucleotide-bound L-head is forced to bind to actin in the R-ARH orientation and hence tends to detach from actin, leading to more frequent foot stomping. This implies that the L-head would be less susceptible to catalytic activation by actin, possibly contributing to some extent to the asymmetry in the kinetic constants between the two heads.

Discussion

In general, the conformational change in myosin upon the hydrolysis of bound ATP to ADP- P_i is the recovery stroke (that is, the post-stroke-to-pre-stroke transition), and the actin-activated P_i release from ADP- P_i reverses the head conformation to produce a forward force²³. However, a forward force (or the strained pre-stroke state) can

also be produced by the direct binding of the ADP-bound L-head to actin. In ATP, the force that is produced in this way after a foot stomp at the L-head drives lever-arm swing. The number of foot stomp events at the L-head per step is around one only for low [ATP]. However, this type of force generation is likely to take place also for high [ATP] under a loaded condition. Moreover, this force generation mechanism suggests that the ADP-bound head has the potential to produce a force repeatedly when an appropriate backward force is applied. This potential was recently indicated in an optical-trap-based experiment on a single myosin V head interacting with actin under a certain backward load³¹. The power stroke is reversed and actin-bound pre- and post-stroke conformations are repeated before one ATPase cycle is completed. In Supplementary Fig. 6, we propose a model for the chemomechanical cycle of myosin V based on previous studies and the new findings obtained here.

As demonstrated here, previously known and unknown behaviours of M5-HMM clearly appear in the molecular movies. Unlike previous methods, high-resolution HS-AFM imaging unselectively provides comprehensive information on the structure and dynamics of a functioning molecule. Thus, the HS-AFM imaging of functioning biomolecules has the potential to transform the fields of structural biology and single-molecule biology.

METHODS SUMMARY

We prepared M5-HMM from chick brain, and prepared partially biotinylated actin filaments and stabilized them with phalloidin¹¹. Biotin-containing lipid bilayers were formed on a mica surface⁸. We immobilized the biotinylated actin filaments on the bilayer surface and then deposited a solution of M5-HMM on the surface. All imaging experiments were performed in the tapping mode using a laboratory-built HS-AFM apparatus^{2,3}.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 June; accepted 24 August 2010.

Published online 10 October 2010.

- Joo, C. *et al.* Advances in single-molecule fluorescence methods for molecular biology. *Annu. Rev. Biochem.* **77**, 51–76 (2008).
- Ando, T., Uchihashi, T. & Fukuma, T. High-speed atomic force microscopy for nano-visualization of dynamic biomolecular processes. *Prog. Surf. Sci.* **83**, 337–437 (2008).
- Ando, T. *et al.* A high-speed atomic force microscope for studying biological macromolecules. *Proc. Natl Acad. Sci. USA* **98**, 12468–12472 (2001).
- Hansma, P. K., Schitter, G., Fantner, G. E. & Prater, C. High-speed atomic force microscopy. *Science* **314**, 601–602 (2006).
- Viani, M. B. *et al.* Probing protein-protein interactions in real time. *Nature Struct. Biol.* **7**, 644–647 (2000).
- Shibata, M. *et al.* High-speed atomic force microscopy shows dynamic molecular processes in photoactivated bacteriorhodopsin. *Nature Nanotechnol.* **5**, 208–212 (2010).
- Yamashita, H. *et al.* Dynamics of bacteriorhodopsin 2D crystal observed by high-speed atomic force microscopy. *J. Struct. Biol.* **167**, 153–158 (2009).
- Yamamoto, D., Uchihashi, T., Kodera, N. & Ando, T. Anisotropic diffusion of point defects in a two-dimensional crystal of streptavidin observed by high-speed atomic force microscopy. *Nanotechnology* **19**, 384009 (2008).
- Sellers, J. R. & Weisman, L. S. in *Myosins: A Superfamily of Molecular Motors, Proteins and Cell Regulation* (ed. Coluccio, L.) 289–324 (Springer, 2008).
- Mehta, A. D. *et al.* Myosin-V is a processive actin-based motor. *Nature* **400**, 590–593 (1999).
- Sakamoto, T., Amitani, I., Yokota, E. & Ando, T. Direct observation of processive movement by individual myosin V molecules. *Biochem. Biophys. Res. Commun.* **272**, 586–590 (2000).
- Yildiz, A. *et al.* Myosin V walks hand-over-hand: single fluorophore imaging with 1.5 nm localization. *Science* **300**, 2061–2065 (2003).
- Forkey, J. N. *et al.* Three-dimensional structural dynamics of myosin V by single-molecule fluorescence polarization. *Nature* **422**, 399–404 (2003).
- Sakamoto, T. *et al.* Direct observation of the mechanochemical coupling in myosin Va during processive movement. *Nature* **455**, 128–132 (2008).
- Walker, M. L. *et al.* Two-headed binding of a processive myosin to F-actin. *Nature* **405**, 804–807 (2000).
- Burgess, S. *et al.* The prepower stroke conformation of myosin V. *J. Cell Biol.* **159**, 983–991 (2002).
- Oke, O. A. *et al.* Influence of lever structure on myosin 5a walking. *Proc. Natl Acad. Sci. USA* **107**, 2509–2514 (2010).
- Baker, J. E. *et al.* Myosin V processivity: multiple kinetic pathways for head-to-head coordination. *Proc. Natl Acad. Sci. USA* **101**, 5542–5546 (2004).

19. Huxley, H. E. The mechanism of muscular contraction. *Science* **164**, 1356–1366 (1969).
20. Hua, W., Chung, J. & Gelles, J. Distinguishing inchworm and hand-over-hand processive kinesin movement by neck rotation measurements. *Science* **295**, 844–848 (2002).
21. Okada, T. *et al.* The diffusive search mechanism of processive myosin class-V motor involves directional steps along actin subunits. *Biochem. Biophys. Res. Commun.* **354**, 379–384 (2007).
22. Wang, F. *et al.* Effect of ADP and ionic strength on the kinetic and motile properties of recombinant mouse myosin V. *J. Biol. Chem.* **275**, 4329–4335 (2000).
23. Geeves, M. A. & Holmes, K. C. Structural mechanism of muscle contraction. *Annu. Rev. Biochem.* **68**, 687–728 (1999).
24. De La Cruz, E. M. *et al.* The kinetic mechanism of myosin V. *Proc. Natl Acad. Sci. USA* **96**, 13726–13731 (1999).
25. Syed, S. *et al.* Adaptability of myosin V studied by simultaneous detection of position and orientation. *EMBO J.* **25**, 1795–1803 (2006).
26. Rosenfeld, S. S. & Sweeney, H. L. A model of myosin V processivity. *J. Biol. Chem.* **279**, 40100–40111 (2004).
27. Veigel, C., Schmitz, S., Wang, F. & Sellers, J. R. Load-dependent kinetics of myosin-V can explain its high processivity. *Nature Cell Biol.* **7**, 861–869 (2005).
28. Oguchi, Y. *et al.* Load-dependent ADP binding to myosins V and VI: implications for subunit coordination and function. *Proc. Natl Acad. Sci. USA* **105**, 7714–7719 (2008).
29. Purcell, T. J., Sweeney, H. L. & Spudich, J. A. A force-dependent state controls the coordination of processive myosin V. *Proc. Natl Acad. Sci. USA* **102**, 13873–13878 (2005).
30. Forgacs, E. *et al.* Kinetics of ADP dissociation from the trail and lead heads of actomyosin V following the power stroke. *J. Biol. Chem.* **283**, 766–773 (2008).
31. Sellers, J. R. & Veigel, C. Direct observation of the myosin-Va power stroke and its reversal. *Nature Struct. Mol. Biol.* **17**, 590–595 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Uchihashi and H. Yamashita for technical assistance and discussion, and J. Sellers for critical reading of the draft and for comments. This work was supported by CREST/JST, Special Coordination Funds for Promoting Science and Technology (Effective Promotion of Joint Research with Industry, Academia, and Government) from JST, a Grant-in-Aid for Basic Research (S) from JSPS, and Knowledge Cluster/MEXT–Japan.

Author Contributions N.K. performed the HS-AFM experiments and data analysis and wrote the first draft of the manuscript. D.Y. and N.K. developed the lipid-bilayer-based assay system. R.I. participated in the sample preparations in the early stage of this study. T.A. and N.K. developed the HS-AFM apparatus. T.A. designed the experiment and prepared the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.A. (tando@kenroku.kanazawa-u.ac.jp).

METHODS

Protein purification and reagents. We prepared M5-HMM from chick brain³². Partially biotinylated (20%) actin filaments were prepared and then stabilized with phalloidin¹¹. Streptavidin, synthetic lipids, apyrase and hexokinase were purchased from Wako Pure Chemical, Avanti Polar Lipids, Sigma-Aldrich and Roche Diagnostics, respectively.

Preparation of samples for AFM observation. We formed biotin-containing lipid bilayers on a mica surface^{8,33}. A typical lipid composition was 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC), 1,2-dipalmitoyl-3-trimethylammonium-propane (DPTAP) and 1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine-*N*-(cap biotinyl) (biotin-cap-DPPE) in a weight ratio of 0.85:0.05:0.1. In some cases, the content of the positively charged lipid DPTAP was reduced. After rinsing the lipid bilayer substrate with buffer A (20 mM imidazole-HCl (pH 7.6), 25 mM KCl, 2 mM MgCl₂, 1 mM EGTA, 5 mM DTT), we deposited a drop (2 μ l) of streptavidin in buffer A (10 nM) on the substrate for 3 min. In some experiments, additional streptavidin molecules (2 μ l, 40 nM) were deposited on the surface for 2 min. After rinsing with buffer A, we deposited a drop (2 μ l) of partially biotinylated actin filaments (1 μ M) in buffer A on the lipid bilayers for 10 min. After rinsing with a solution containing either 0.1 μ M to 1 mM ATP, 0.1 μ M to 1 mM ADP (1 U ml⁻¹ hexokinase and 10 mM glucose were added in the case of 1 mM ADP to remove contaminating ATP) or 1 U ml⁻¹ apyrase (to ensure the nucleotide-free condition) in buffer A, we deposited a drop (2 μ l) of the same solution plus M5-HMM (0.1 to 1 nM) on the lipid bilayers for 3 min. Finally, the sample was attached to the scanning stage of a HS-AFM apparatus and immersed in the same solution (~60 μ l) without M5-HMM. When the lipid bilayers did not contain the positively charged lipid DPTAP, we further added M5-HMM to the observation solution.

HS-AFM apparatus and cantilevers. We used a laboratory-built tapping-mode HS-AFM apparatus^{2,3} together with small cantilevers designed for HS-AFM³⁴ (spring constant, 0.1–0.2 N m⁻¹; resonant frequency in water, 0.8–1.2 MHz; quality factor in water, ~2). The probe tip was grown on the tip of a cantilever by electron-beam deposition and was further sharpened by argon plasma etching². In the best case, a tip apex radius of ~4 nm was achieved. To achieve a small tip-sample loading force, the free-oscillation peak-to-peak amplitude of the cantilever (A_0) was set to 1.5–2.5 nm and the amplitude set point was set at more than 0.9 A_0 . We note that the mechanical quantity that affects the sample is not the force itself but the force impulse, that is, the product of force and the time over which the force acts. In tapping-mode HS-AFM, the time of force action is short (~100 ns) and, therefore, a relatively large peak force (~20 pN) would not affect the sample significantly. All the imaging experiments could be routinely conducted without difficulty except for the occasional deterioration of image quality owing to tip smear or cracking, which was difficult to control. The tip smearing was caused by the attachment of molecules contained in the sample, but they were sometimes removed during imaging. The cracking of the sharp tip yields a double image of every object, as shown in Supplementary Movie 6.

HS-AFM imaging and processive run. All observations were performed at room temperature (24–26 °C). To facilitate rapid imaging of moving M5-HMM molecules on actin, we chose sample areas in which actin filaments were aligned approximately parallel to the x direction. Myosin V walks approximately straight along actin filaments as the two heads span the filament's helical pseudorepeat of ~36 nm (refs 10, 12–15). This property and the high-precision feedback system of the instrument^{2,35} facilitate the continuous imaging of walking M5-HMM molecules weakly adsorbed sideways onto the DPTAP-containing substrate surface. In the absence of DPTAP in the substrate, M5-HMM molecules often walked on the top surface of actin filaments. In this case, it was difficult to acquire high-resolution images because AFM is inefficient for visualizing molecules not supported on a surface.

High-resolution imaging of moving M5-HMM. To obtain high-resolution images of M5-HMM molecules translocating along actin in 0.1–2 μ M ATP, the scan area was narrowed (typically to about 130 \times 65 nm²) and the number of pixels was optimized (80 \times 40 or 80 \times 50) such that the phase delay in tracing the sample surface was minimized and the imaging rate (146.7 ms per frame) was high enough to capture the moving molecules. When long processive runs (~15 steps) in 0.1–2 μ M ATP were observed, the molecules were tracked by shifting the narrow scan area manually (mainly in the x direction). The tracking sometimes could not be continued when the target molecule collided with a molecule walking ahead, when a portion of discontinuous substrate surface was encountered or when the manual tracking by an operator was not appropriately made. The occasional failure of continual tracking is not due to the tip-sample interaction.

Velocity measurements. In the experiment to determine the translocation velocity of M5-HMM for various [ATP] values, we fixed the scan areas (400 \times 125 nm²) and

captured images at 146.3 ms per frame. To facilitate rapid imaging of this relatively wide scan area, the number of pixels was reduced to 80 \times 25 (thus, the scan velocity in the x direction is only slightly higher than that in high-resolution imaging mentioned above). The image quality was generally lowered but was high enough to identify M5-HMM molecules on actin. When the velocity of a translocating molecule was high, some of the dwell events could not be detected. In this case, we determined it using at least four different dwell positions where the molecule being tracked was captured on video.

Imaging for detecting foot stomp and L-head conformations. In the experiment to analyse foot stomping and L-head conformations in ADP and under the nucleotide-free conditions, we carried out the imaging at 333.2 ms per frame using fixed imaging areas (typically 90 \times 90 nm² with 80 \times 80 pixels). In the experiment to analyse foot stomping and L-head conformations in ATP, the imaging was carried out at 146.7 ms per frame using two scan modes. In one mode, the imaging areas were fixed (typically at 130 \times 65 nm² with 80 \times 40 pixels). M5-HMM molecules that entered the imaging areas were captured on video. In the other mode, a translocating M5-HMM molecule was tracked by shifting the scan area (typically 130 \times 65 nm² with 80 \times 40 pixels).

Image analysis. We applied to each AFM image a Gaussian filter for removing spike noises and a flattening filter for removing the substrate-tilt effect. Each position of the motor domain of M5-HMM was determined by calculating the centre of mass to measure its displacement. The following events were considered as a foot stomp: (1) an entire head disappears and then reappears at approximately the same position; (2) either motor domain of a two-headed bound M5-HMM molecule is translocated by around ± 5 nm along the actin filament; and (3) a motor domain disappears while its neck domain is visible (that is, the head detaches while scanning the region where the motor domain was attached in the previous frame scan, and then rebinds to actin). Moreover, when an entire T-head disappeared in ADP and under the nucleotide-free condition, the molecule sometimes stepped forward. This event was also counted as a foot stomp. We note that a foot stomp was also observed even when the substrate surface did not contain DPTAP, although we did not systematically analyse this. When the L-head is sharply bent, the neck-neck junction becomes closer to the bound actin filament (and, hence, the space surrounded by the two heads and actin becomes narrower). We identified the sharply bent conformation by visual observation of these features. For the quantitative analysis of foot stomp events and L-head conformations in ATP, we analysed only images within which both heads were fully located before and after a step. Therefore, we observed more steps than are shown in Fig. 3a, b.

Kinetic analysis of conformational change at L-head. We analysed the proportion of the straight L-head conformation detected (r) as a function of [ADP] (Fig. 3b) to estimate the ADP dissociation constant for the L-head (K_d) using the relationship

$$r([\text{ADP}]) = \frac{r_D[\text{ADP}] + r_{\text{NF}}K_d}{[\text{ADP}] + K_d}$$

where $r_D = 0.98$ and $r_{\text{NF}} = 0.36$ are the proportions of straight L-head detected in a saturating amount of ADP and under the nucleotide-free condition, respectively. Considering the reaction scheme (Fig. 3c, inset), the average lifetime of the straight L-head, $\langle\tau_{\text{ST}}\rangle$, is expressed approximately as

$$\langle\tau_{\text{ST}}\rangle = n_{\text{NF}} \left(k_{\text{S-B}} + \frac{k_- k_+ [\text{ADP}]}{k_- + k_+ [\text{ADP}]} \right)^{-1} + n_{\text{ADP}} \left(k_- + \frac{k_- k_{\text{S-B}}}{k_- + k_{\text{S-B}}} \right)^{-1} \quad (1)$$

where $k_{\text{S-B}}$ represents the spontaneous straight-to-bent transition rate for the nucleotide-free L-head; n_{NF} and n_{ADP} are the fractions of nucleotide-free L-head and ADP-bound L-head for a given [ADP] value, respectively, which are calculated using $K_d = 0.075 \mu\text{M}$; and k_- and k_+ are the ADP dissociation rate constant and ADP association rate constant at the L-head, respectively. Using equation (1), the observed $\langle\tau_{\text{ST}}\rangle$ as a function of [ADP] (Fig. 3c) was analysed to estimate $k_{\text{S-B}}$, k_- and k_+ .

- Koide, H. *et al.* Identification of the single specific IQ motif of myosin V from which calmodulin dissociates in the presence of Ca²⁺. *Biochemistry* **45**, 11598–11604 (2006).
- Yamamoto, D. *et al.* High-speed atomic force microscopy techniques for observing dynamic biomolecular processes. *Methods Enzymol.* **475**, 541–564 (2010).
- Kitazawa, M., Shiotani, K. & Toda, A. Batch fabrication of sharpened silicon nitride tips. *Jpn. J. Appl. Phys.* **42**, 4844–4847 (2003).
- Kodera, N., Sakashita, M. & Ando, T. Dynamic proportional-integral-differential controller for high-speed atomic force microscopy. *Rev. Sci. Instrum.* **77**, 083704 (2006).

Formation of the black-hole binary M33 X-7 through mass exchange in a tight massive system

Francesca Valsecchi¹, Evert Glebbeek², Will M. Farr¹, Tassos Fragos^{1,5}, Bart Willems¹, Jerome A. Orosz³, Jifeng Liu^{4,5} & Vassiliki Kalogera¹

The X-ray source M33 X-7 in the nearby galaxy Messier 33 is among the most massive X-ray binary stellar systems known, hosting a rapidly spinning, $15.65M_{\odot}$ black hole orbiting an underluminous, $70M_{\odot}$ main-sequence companion in a slightly eccentric 3.45-day orbit^{1,2} (M_{\odot} , solar mass). Although post-main-sequence mass transfer explains the masses and tight orbit³, it leaves unexplained the observed X-ray luminosity, the star's underluminosity, the black hole's spin and the orbital eccentricity. A common envelope phase¹, or rotational mixing⁴, could explain the orbit, but the former would lead to a merger and the latter to an overluminous companion. A merger would also ensue if mass transfer to the black hole were invoked for its spin-up⁵. Here we report simulations of evolutionary tracks which reveal that if M33 X-7 started as a primary body of $85M_{\odot}$ – $99M_{\odot}$ and a secondary body of $28M_{\odot}$ – $32M_{\odot}$, in a 2.8–3.1-d orbit, its observed properties can be consistently explained. In this model, the main-sequence primary transfers part of its envelope to the secondary and loses the rest in a wind; it ends its life as a $\sim 16M_{\odot}$ helium star with an iron–nickel core that collapses to a black hole (with or without an accompanying supernova). The release of binding energy, and possibly collapse asymmetries, 'kick' the nascent black hole into an eccentric orbit. Wind accretion explains the X-ray luminosity, and the black-hole spin can be natal.

M33 X-7 has been identified as an evolutionary challenge, given the massive components and the tight orbit of the system relative to the size of the large hydrogen-rich companion. Four mechanisms have been proposed to explain the formation of M33 X-7, but none of them has

addressed nor can simultaneously explain all its observed properties (Tables 1 and 2). We performed detailed binary evolution calculations to explore possible evolutionary tracks. Given the spatial metallicity gradient of Messier 33⁶ (M33) we assume a metallicity of 50% of the solar value for all our models.

To illustrate the evolutionary history of M33 X-7 clearly, we show the results for one of the successful evolutionary sequences in Fig. 1. The binary progenitor comprises a primary of $\sim 97M_{\odot}$ (the progenitor of the black hole) and a secondary of $\sim 32M_{\odot}$ (the progenitor of the black-hole companion) in an orbit of ~ 2.9 d. During the first ~ 1.8 Myr, the evolution is driven by mass loss due to stellar winds, causing a decrease in the gravitational attraction between the components and the increase of the orbit to ~ 3.25 d. The more massive primary evolves faster than the secondary, growing in size to accommodate the energy produced by fusing hydrogen into helium at its centre. Eventually, while still on the main sequence, it expands and begins mass transfer to the secondary when it enters the sphere of influence of the secondary's gravitational field (through Roche-lobe overflow). This stronger mode of mass loss brings the primary out of thermal equilibrium; in response, the star shrinks, recovering its thermal equilibrium while always maintaining hydrostatic equilibrium and, hence, dynamical stability. During the first few tens of thousands of years of mass transfer, the orbital period decreases because the more massive primary is transferring mass to the less massive secondary. When the secondary accretes enough matter to become the more massive component, the orbit starts expanding⁷. The primary transfers most of its hydrogen-rich envelope and becomes a Wolf–Rayet star, and the strong Wolf–Rayet wind ($\sim (2\text{--}3) \times 10^{-5}M_{\odot} \text{ yr}^{-1}$) removes much

Table 1 | Observed parameters for M33 X-7 for the reference distance of 840 ± 20 kpc adopted by the discovery team¹

Parameter	Value
M_{BH}	$15.65M_{\odot} \pm 1.45M_{\odot}$
M_2	$70.0M_{\odot} \pm 6.9M_{\odot}$
Spectral type	O7III to O8III
T_{eff}	$35,000 \pm 1,000$ K
$\log(L_2/L_{\odot})$	5.72 ± 0.07
P	3.45301 ± 0.00002 d
e	0.0185 ± 0.0077
i	$74.6 \pm 1.0^{\circ}$
L_X	$(0.13\text{--}2.49) \times 10^{38} \text{ erg s}^{-1}$
a^*	0.84 ± 0.05

The black-hole mass (M_{BH}), the mass (M_2), spectral type, effective temperature (T_{eff}) and luminosity (L_2) of its companion, the orbital eccentricity (e) and the inclination (i) are as reported in ref. 1. The orbital period (P) is as measured in ref. 2. The dimensionless spin parameter of the black hole (a^*) was determined in ref. 9 on the basis of the X-ray continuum fitting method^{16–18}. The X-ray luminosity (L_X) is derived from observations reported in refs 1, 2, 9, 19–21. To account for variations in the X-ray flux over different observations, we consider the lowest and highest reported values, after we rescale each L_X to an M33 distance of 840 ± 20 kpc. If the full distance range of 750–1,017 kpc is adopted, using the ELC code of ref. 22 we calculate the masses to be $55M_{\odot}$ – $103M_{\odot}$ and $13.5M_{\odot}$ – $20M_{\odot}$ for the star and the black hole, respectively, and the inclination is 77° – 71° . The logarithmic luminosity in solar units is then 5.62–5.89 (ref. 1). For each distance, the mass of the star can be derived from $M_2 = -75.94 + 0.17d$ and the inclination from $i = 93.69 - 0.02d$, where d is in kiloparsecs, M_2 is in solar masses and i is in degrees. For each M_2 , the corresponding black hole's mass in solar masses can be calculated from $M_{\text{BH}} = 6.19 + 0.13M_2$, and its spin can be calculated from $a^* = 0.31 + 0.049M_{\text{BH}} - 0.001M_{\text{BH}}^2$. The rescaled X-ray luminosity ranges from $\sim 1 \times 10^{37}$ to $\sim 3 \times 10^{38} \text{ erg s}^{-1}$.

Table 2 | Models suggested for the formation of M33 X-7

Model	M_{BH}	M_2	T_{eff}	$\log(L_2)$	P	e	L_X	a^*
Ref. 1	X	O	O	O	X	O	X	O
Ref. 3	✓	✓	O	O	✓	O	O	O
Ref. 4	X	X	O	O	X	O	O	O
Ref. 5	O	O	O	O	O	O	O	✓

The various parameters are described in Table 1. The symbol '✓' means a parameter has been addressed and explained, 'X' means a parameter has been addressed but not explained and 'O' means a parameter has not been addressed at all. To explain the tight orbit, ref. 1 suggests that the progenitors underwent a common envelope phase during which the primary expanded to the point of engulfing its companion. Such a phase is known to produce tight systems because energy can be transferred from the orbit to the common envelope, leading to a reduction in the binary separation and ejection of the envelope²³. To form the observed black hole, ref. 1 requires that the common envelope begins after helium core burning in the primary is complete. However, a common envelope in the case of M33 X-7 would probably evolve into a merger, because massive-star envelopes are tightly bound²⁴. Furthermore, for this model to succeed an unrealistically low stellar wind would be required. Ref. 3 suggests a phase of conservative mass transfer from the black-hole progenitor to the companion that sets in when the primary leaves the main sequence, but this model only explains the observed masses and orbital period, and fails to address the remaining observations. Ref. 4 proposes rotationally induced mixing as a way to keep massive stars from expanding significantly during the main sequence, preventing mass transfer or a merger until the primary becomes a Wolf–Rayet star. However, this evolutionary channel increases the star's luminosity above that of standard models, which contradicts the observed underluminosity of the star in M33 X-7. Ref. 5 explains the observed black hole's spin in terms of a past Roche-lobe overflow phase from the star to the black hole. However, given the extreme mass ratio between the components, such a phase would have evolved into a merger.

¹Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA) and Department of Physics and Astronomy, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, USA. ²Department of Physics and Astronomy, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4M1, Canada. ³Department of Astronomy, San Diego State University, 5500 Campanile Drive, San Diego, California 92182-1221, USA. ⁴National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China. ⁵Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.

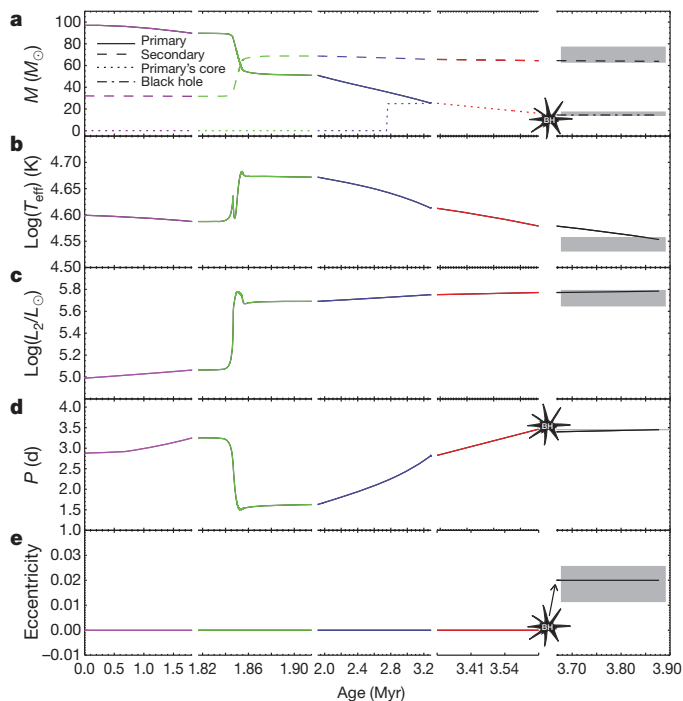


Figure 1 | Evolution of the orbital and stellar parameters of M33 X-7.

a, Masses (M); **b**, secondary's surface temperature ($\log(T_{\text{eff}})$); **c**, secondary's luminosity ($\log(L_2)$); **d**, orbital period (P); **e**, eccentricity (e). The different evolutionary stages are highlighted with different colours: the beginning of the main sequence, purple; the mass transfer phase, green; the end of the main sequence, blue; the core helium burning phase for the black-hole progenitor, red; the phase between black-hole formation and the present time, black. (Note the non-uniform x axis.) The grey shaded areas represent the observational constraints as reported in Table 1. The sequence comprises an $M_1 \approx 97M_{\odot}$ primary and an $M_2 \approx 32M_{\odot}$ secondary in a ~ 2.9 -d orbit. At the onset of the mass transfer phase (purple/green), $M_1 \approx 89.7M_{\odot}$, $M_2 \approx 31.7M_{\odot}$ and $P \approx 3.25$ d. During the mass transfer phase (green), the primary transfers conservatively $\sim 37M_{\odot}$ to the companion (see Supplementary Information for details about the mass transfer). When the system detaches (green/blue), $M_1 \approx 51M_{\odot}$, $M_2 \approx 68.7M_{\odot}$ and $P \approx 1.6$ d. When the primary leaves the main sequence (blue/red), $M_1 \approx 25.2M_{\odot}$, $M_2 \approx 65.6M_{\odot}$ and $P \approx 2.8$ d. Considering evolutionary models of helium stars (Supplementary Information), we find that a helium star of $\sim 25M_{\odot}$ burns helium at its centre for ~ 0.38 Myr and loses $\sim 9.1M_{\odot}$ to its Wolf-Rayet wind. During this time (red), the secondary loses $\sim 1.1M_{\odot}$. Before black-hole formation (red/black), $M_1 \approx 16M_{\odot}$, $M_2 \approx 64.5M_{\odot}$ and $P \approx 3.46$ d. For the case shown, at the primary's collapse a kick of 120 km s^{-1} is imparted to the newly formed black hole (see Supplementary Information for the allowed kicks).

of the remaining envelope, eventually interrupting the mass transfer. During the ~ 99 kyr of conservative mass transfer, the original $32M_{\odot}$ secondary becomes a massive, $\sim 69M_{\odot}$ O-type star and the primary becomes a $\sim 51M_{\odot}$ Wolf-Rayet star. Once the Wolf-Rayet wind begins and the mass transfer is interrupted, the wind blows away the remaining primary's envelope to expose the $\sim 25M_{\odot}$ helium core. This mass loss drives further orbital expansion until the primary leaves the main sequence and throughout the phase of core helium burning. At the same time, the secondary, which is now more massive, is losing mass to its own O-star wind at a lower rate ($\sim 10^{-6}M_{\odot} \text{ yr}^{-1}$). At this time, the orbit of the binary is circular and the spin period of each star is expected to be synchronized with the orbital period. The synchronization is due to exchange of angular momentum between the stars and their orbit, caused by tidal interaction.

The final stages of the primary's life during and after carbon burning are too short (~ 60 yr for an initially $\sim 25M_{\odot}$ helium star⁸) to change the stellar and orbital parameters significantly. At the end of the primary's life, after ~ 3.7 Myr, M33 X-7 comprises a $\sim 16M_{\odot}$, evolved Wolf-Rayet star with an iron-nickel core and a $\sim 64.5M_{\odot}$ O-star companion in a

~ 3.5 -d orbit. Unable to support itself through further nuclear fusion, this massive, helium-rich star collapses into a black hole and a small fraction of the rest-mass energy (10%) is released as the black hole's gravitational energy. Additionally, asymmetries in the collapse and associated neutrino emission may impart an instantaneous linear momentum recoil (kick) to the newly born black hole, even if there is no baryonic mass ejection at collapse. Both of these effects modify the orbital configuration, inducing an eccentricity and slightly decreasing the orbital period to ~ 3.4 d. The release in binding energy leads to an increase in the orbital separation and the kick imparted to the black hole acts to decrease it. For the ~ 0.2 Myr after the formation of the black hole, the binary evolution is driven by mass loss due to the secondary's stellar wind, causing the orbit to expand further, to the currently observed value. The fraction of this stellar wind attracted and accreted by the black hole is too small to influence the orbital evolution significantly, but is adequate to explain the X-ray luminosity observed. Today, after ~ 3.9 Myr, M33 X-7 comprises a black hole of $\sim 14.4M_{\odot}$ and an underluminous O star of $\sim 64M_{\odot}$ orbiting around their common centre of mass in a slightly eccentric, ~ 3.45 -d orbit.

Our model is consistent with a natal nature of the black hole's observationally inferred high spin⁹. In fact, although it has been suggested that such a high spin is the result of a mass transfer from the companion star to the black hole through Roche-lobe overflow⁵, given how much more massive the companion is relative to the black hole, such a phase could not have occurred: the mass transfer would have been dynamically unstable and would have rapidly evolved into a merger of the binary components. Wind accretion is too weak to spin up the black hole to the current value if it had been formed spinning much slower. In our model, when the black-hole progenitor leaves the main sequence, the spins of the stars are expected to be synchronized with the orbit, and, assuming rigid-body rotation on the main sequence, the inner parts of the primary's core carry enough angular momentum to explain the currently observed black-hole spin (the inner $15.5M_{\odot}$ of the core carry $\sim 5.2 \times 10^{51} \text{ g cm}^2 \text{ s}^{-1}$ of angular momentum, and $\sim 1.8 \times 10^{51} \text{ g cm}^2 \text{ s}^{-1}$ is needed to explain the currently observed spin). At the end of the main sequence, the inner core of the star is expected to decouple rotationally from the outer envelope and approximately retain the angular momentum of the central layers (see table 4 in ref. 10).

We explore binary evolutionary sequences with different combinations of initial masses and orbital periods. We select as 'successful' sequences those that eventually match all observed properties within 1σ errors. All these sequences follow a path qualitatively very similar to the specific example described in detail here. The progenitors are constrained to host $96M_{\odot}$ – $99M_{\odot}$ primaries and $32M_{\odot} \pm 1M_{\odot}$ secondaries in orbits with initial periods of 2.8–2.9 d. The apparent puzzling underluminosity of the black-hole companion is due to two factors: the orientation of the system with respect to our line of sight and associated projection effects reduce the star's measured luminosity (accounting for 87% of the underluminosity), and the secondary was not formed as a $\sim 63M_{\odot}$ – $65M_{\odot}$ star but instead accreted much of its mass (accounting for 13% of the underluminosity) from the black-hole progenitor (Supplementary Information).

We note that the uncertainty in the distance to M33, ~ 840 kpc, is greater than ± 20 kpc, and that some of the observed system properties vary if a different distance to the system is considered (Table 1). Various studies in the literature have reported the distance to be in the range 750–1,017 kpc (Supplementary Information). Considering this full range, the progenitors are constrained to host primaries of $85M_{\odot}$ – $99M_{\odot}$, secondaries of $28M_{\odot}$ – $32M_{\odot}$ and initial orbital periods of 2.8–3.1 d (Fig. 2).

The different phases we have described for the evolutionary past of M33 X-7 have been observed in a variety of other binary systems. For example, the super star cluster R136 in the Large Magellanic Cloud hosts two non-interacting O stars of $\sim 57M_{\odot}$ and $\sim 23M_{\odot}$ orbiting around each other every ~ 3.4 d (LMC R136-38, ref. 11). The eclipsing

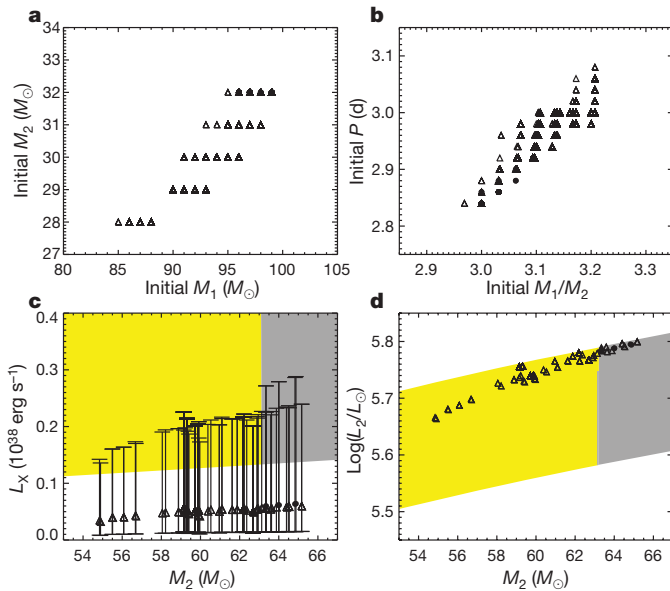


Figure 2 | Progenitor properties and current luminosity. Data represented by circles and triangles are the results of detailed binary star evolution calculations for all successful sequences for M33 distances of 840 ± 20 kpc and $750\text{--}1,017$ kpc, respectively. **a**, Possible masses for the progenitors; **b**, possible initial orbital periods as a function of the mass ratio between the primary and the secondary; **c**, **d**, black-hole X-ray luminosity (**c**) and secondary's luminosity (**d**) as functions of the secondary's mass at present. The grey and yellow shaded areas represent the observational constraints for M33 distances of 840 ± 20 kpc and $750\text{--}1,017$ kpc, respectively. L_X is calculated according to the Bondi–Hoyle accretion model²⁵. The error bars are derived from the uncertainties in the stellar wind parameters (Supplementary Information) and depict the highest and lowest L_X values; they do not represent statistical 1σ errors. The observational constraints on L_2 are calculated given the dependence of M_2 and L_2 on the distance as described in the caption of Table 1, and accounting for uncertainties in the star's effective temperature, reddening and apparent magnitude calculated using the ELC code²². According to our model, secondaries at present more massive than $\sim 65M_\odot$ fail to explain the observed luminosity. Some of the data points are omitted for clarity.

binary LMC-SC1-105 comprises a $\sim 31M_\odot$ O star that is transferring mass to its $\sim 13M_\odot$ companion¹². In the system WR46, a $\sim 51M_\odot$ Wolf–Rayet star orbits a $\sim 60M_\odot$ O-star companion every ~ 6 d. In particular, the various discoveries of Wolf–Rayet stars with O-star companions¹³ validate our theoretical model, as they resemble the configuration of M33 X-7 less than 2 Myr ago. Several of these binaries have been observed with orbital periods of a few days and massive components^{14,15} (up to $\sim 80M_\odot$).

Received 22 December 2009; accepted 26 August 2010.

Published online 20 October 2010.

- Orosz, J. A. *et al.* A 15.65-solar-mass black hole in an eclipsing binary in the nearby spiral galaxy M 33. *Nature* **449**, 872–875 (2007).
- Pietsch, W. *et al.* M33 X-7: ChASeM33 reveals the first eclipsing black hole X-ray binary. *Astrophys. J.* **646**, 420–428 (2006).
- Abubekrov, M. K., Antokhina, E. A., Bogomazov, A. I. & Cherepashchuk, A. M. The mass of the black hole in the X-ray binary M33 X-7 and the evolutionary status of M33 X-7 and IC 10 X-1. *Astron. Rep.* **53**, 232–242 (2009).

- de Mink, S. E. *et al.* Rotational mixing in massive binaries. Detached short-period systems. *Astron. Astrophys.* **497**, 243–253 (2009).
- Moreno Méndez, E., Brown, G. E., Lee, C. & Park, I. H. The case for hypercritical accretion in M33 X-7. *Astrophys. J.* **689**, L9–L12 (2008).
- Vivian, U. *et al.* A new distance to M33 using blue supergiants and the FGLR method. *Astrophys. J.* **704**, 1120–1134 (2009).
- Verbunt, F. Origin and evolution of X-ray binaries and binary radio pulsars. *Annu. Rev. Astron. Astrophys.* **31**, 93–127 (1993).
- Tauris, T. M. & van den Heuvel, E. in *Compact Stellar X-Ray Sources* (eds Lewin, W. & van der Klis, M.) 623–665 (Cambridge Univ. Press, 2006).
- Liu, J., McClintock, J. E., Narayan, R., Davis, S. W. & Orosz, J. A. Erratum: “Precise measurement of the spin parameter of the stellar-mass black hole M33 X-7”. *Astrophys. J.* **719**, L109 (2010).
- Hirschi, R., Meynet, G. & Maeder, A. Stellar evolution with rotation. XIII. Predicted GRB rates at various Z. *Astron. Astrophys.* **443**, 581–591 (2005).
- Massey, P., Penny, L. R. & Vukovich, J. Orbits of four very massive binaries in the R136 cluster. *Astrophys. J.* **565**, 982–993 (2002).
- Bonanos, A. Z. Toward an accurate determination of parameters for very massive stars: the eclipsing binary LMC-SC1–105. *Astrophys. J.* **691**, 407–417 (2009).
- van der Hucht, K. A. The Vllth catalogue of galactic Wolf–Rayet stars. *N. Astron. Rev.* **45**, 135–232 (2001).
- Rauw, G. *et al.* WR 20a: a massive cornerstone binary system comprising two extreme early-type stars. *Astron. Astrophys.* **420**, L9–L13 (2004).
- Bonanos, A. Z. *et al.* WR 20a is an eclipsing binary: accurate determination of parameters for an extremely massive Wolf–Rayet system. *Astrophys. J.* **611**, L33–L36 (2004).
- Zhang, S. N., Cui, W. & Chen, W. Black hole spin in X-ray binaries: observational consequences. *Astrophys. J.* **482**, L155–L158 (1997).
- Li, L., Zimmerman, E. R., Narayan, R. & McClintock, J. E. Multitemperature blackbody spectrum of a thin accretion disk around a Kerr black hole: model computations and comparison with observations. *Astrophys. J. Suppl. Ser.* **157**, 335–370 (2005).
- McClintock, J. E. *et al.* The spin of the near-extreme Kerr black hole GRS 1915+105. *Astrophys. J.* **652**, 518–539 (2006).
- Parmar, A. N. *et al.* BeppoSAX spectroscopy of the luminous X-ray sources in M 33. *Astron. Astrophys.* **368**, 420–430 (2001).
- Pietsch, W. *et al.* The eclipsing massive X-ray binary M 33 X-7: new X-ray observations and optical identification. *Astron. Astrophys.* **413**, 879–887 (2004).
- Shporer, A., Hartman, J., Mazeh, T. & Pietsch, W. Photometric analysis of the optical counterpart of the black hole HMXB M 33 X-7. *Astron. Astrophys.* **462**, 1091–1095 (2007).
- Orosz, J. A. & Hauschildt, P. H. The use of the NextGen model atmospheres for cool giants in a light curve synthesis code. *Astron. Astrophys.* **364**, 265–281 (2000).
- Taam, R. E. & Sandquist, E. L. Common envelope evolution of massive binary stars. *Annu. Rev. Astron. Astrophys.* **38**, 113–141 (2000).
- Podsiadlowski, P., Rappaport, S. & Han, Z. On the formation and evolution of black hole binaries. *Mon. Not. R. Astron. Soc.* **341**, 385–404 (2003).
- Bondi, H. & Hoyle, F. On the mechanism of accretion by stars. *Mon. Not. R. Astron. Soc.* **104**, 273–282 (1944).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank N. Ivanova, A. Heger, A. Cantrell and C. Bailyn for discussions during the development of this project.

Author Contributions V.K. and B.W. designed the study. F.V. led the project, performed the single and binary star evolution calculations, and developed the code to perform the orbital evolution after black-hole formation. V.K. and B.W. collaborated with F.V. in each step of the project. E.G. maintained, updated and extended the stellar evolution code used, and collaborated with F.V. in performing the calculations. W.M.F. determined the correction to the star's luminosity and surface temperature due to the inclination of the system. T.F. led the theoretical analysis of the black-hole spin. J.A.O. performed the analysis of the observational data on M33 X-7 using the ELC code for the full distance uncertainty. J.L. recalculated the black hole's spin at different distances. All authors discussed the results and made substantial contributions to the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.V. (francesca@u.northwestern.edu).

Holographic three-dimensional telepresence using large-area photorefractive polymer

P.-A. Blanche¹, A. Bablumian¹, R. Voorakaranam¹, C. Christenson¹, W. Lin², T. Gu², D. Flores², P. Wang², W.-Y. Hsieh², M. Kathaperumal², B. Rachwal², O. Siddiqui², J. Thomas¹, R. A. Norwood¹, M. Yamamoto² & N. Peyghambarian¹

Holography is a technique that is used to display objects or scenes in three dimensions. Such three-dimensional (3D) images, or holograms, can be seen with the unassisted eye and are very similar to how humans see the actual environment surrounding them. The concept of 3D telepresence, a real-time dynamic hologram depicting a scene occurring in a different location, has attracted considerable public interest since it was depicted in the original *Star Wars* film in 1977. However, the lack of sufficient computational power to produce realistic computer-generated holograms¹ and the absence of large-area and dynamically updatable holographic recording media² have prevented realization of the concept. Here we use a holographic stereographic technique³ and a photorefractive polymer material as the recording medium⁴ to demonstrate a holographic display that can refresh images every two seconds. A 50 Hz nanosecond pulsed laser is used to write the holographic pixels⁵. Multicoloured holographic 3D images are produced by using angular multiplexing, and the full parallax display employs spatial multiplexing. 3D telepresence is demonstrated by taking multiple images from one location and transmitting the information via Ethernet to another location where the hologram is printed with the quasi-real-time dynamic 3D display. Further improvements could bring applications in telemedicine, prototyping, advertising, updatable 3D maps and entertainment.

3D display technology is attracting much public attention; events include the recent release of 3D films such as *Avatar*, the 2008 US election-night 'hologram' reporter interviews from CNN (<http://www.cnn.com/2008/TECH/11/06/hologram.yellin/index.html>), and the demonstration of 3D televisions by some manufacturers (<http://www.3dsource.com/>). As dramatic as these effects are, the technology used is based on polarization stereoscopy (the technique currently used in cinemas and television), digital image fusion (in the case of CNN), or 2D semitransparent screens (for musion; <http://www.musion.co.uk/>). As such, they have little to do with holography, which is the reproduction of the amplitude and phase of light by diffraction⁶. Nevertheless, these examples show the great enthusiasm that the public, media and industry share about 3D image rendering, and for a good reason: the human physiology has adapted to observe its environment in three dimensions.

Holography, with its ability to reconstitute both the intensity and wave front information of a scene, allows the observer to perceive the light as it would have been scattered by the real object itself⁷. Furthermore, there is no need for any special eyewear to be worn by the observer. It has been shown that holograms can be computer generated⁸. Unfortunately, the amount of information needed to produce a high quality hologram is so large that making a real-time video-rate display has been limited by either size or resolution^{9–11}. To overcome those major issues, different solutions have been tested, such as pupil tracking¹², the use of a holographic diffuser screen^{13,14} or the synthesis of a hologram from real objects¹⁵. It has also been shown that the holographic stereographic technique¹⁶ (also referred to as integral holography), using the diffraction of light to reproduce both parallax

and occlusion clues (but not reproducing the phase), can be used to ease data management. Compared to normal stereograms or anaglyphs, holographic stereography does not require the viewer to use eyewear to perceive the 3D effect. The technique reconstitutes multiple perspectives that the observer can experience by looking at the screen from different angles. Applied to permanent holographic media such as silver halide films or photopolymers, holographic stereography is capable of providing excellent resolution and depth reproduction on large-scale 3D static images (<http://www.zebraimaging.com>); but dynamic updating capability has been missing until now.

Photorefractive inorganic crystals have been used in the past to demonstrate refreshable holographic displays^{17,18}. However, such systems suffer from the disadvantage that crystalline materials are not scalable in size owing to their laborious growth process, and thus are not well suited for display applications. Our group has introduced an updatable holographic 3D display based on the holographic stereographic technique using photorefractive polymeric materials¹⁹. Although this was an important step towards dynamic holography, the system had several limitations, including being monochromatic and having a low refresh rate (more than 4 min per image)^{20,21}; the rapid updating needed for video rates or 3D telepresence was not possible.

Here we describe a new holographic 3D display based on a novel photorefractive material capable of refreshing images every two seconds, making it the first to achieve a speed that can be described as quasi-real-time. The system is based on a pulsed laser holographic recording system and a new sensitized photorefractive polymer with remarkable holographic properties. Each holographic pixel, known as a 'hogel'⁵, is written with a single nanosecond laser pulse. As opposed to 2D pixels, hogels contain 3D information from various perspectives. We also demonstrate multi-colour capability using angular multiplexing and

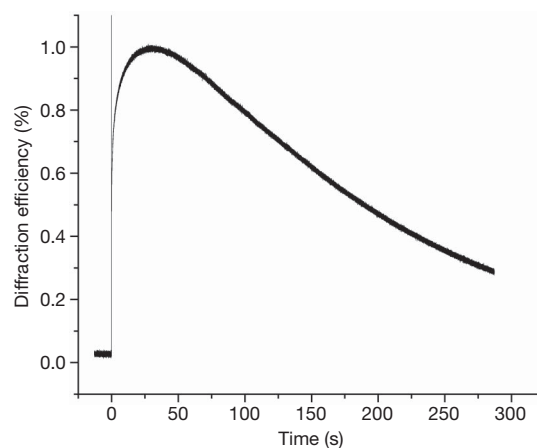


Figure 1 | Example of diffraction efficiency dynamics under single nanosecond pulse writing. The pulsed energy at the sample location was 650 mJ cm^{-2} (sum of both beams). Applied voltage to the photorefractive device was 7 kV and kept constant during the whole measurement.

¹College of Optical Sciences, The University of Arizona, Tucson, Arizona 85721, USA. ²Nitto Denko Technical Corporation, Oceanside, California 92054, USA.



Figure 2 | Image from a hologram recorded with the pulsed system. Panels a, b, c respectively show images observed by the camera when pointed to the left, straight ahead, and to the right. This shows the 3D nature of the image (a

model aeroplane) by demonstrating parallax and occlusions. Supplementary Movie 1 shows the recording at 50 Hz and the display of the hologram.

full parallax at the same speed as horizontal-parallax-only (HPO). 3D telepresence is demonstrated with the HPO system, whereby 2D images were taken at multiple angles from one place and sent to another location using the Ethernet communication protocol, and then printed with the holographic set-up. To the best of our knowledge this is the first demonstration of holographic 3D telepresence.

Our earlier rewritable 3D display¹⁹ used a continuous-wave frequency-doubled Nd:YAG laser as the recording source. The recording time was about one second per hogel, and the scanning system (used to shift from one hogel position to the next) needed to be stopped and damped each time to avoid vibration, resulting in a time-consuming process. As a result, the overall recording time for a 4 inch \times 4 inch hologram consisting of 120 hogels was of the order of 3 min. An erase time of 1 min was needed before refreshing the image. The system suffered from high sensitivity to ambient noise (such as vibration and air turbulence) and to thermal expansion, requiring a fully enclosed air damped optical table.

In this work we have used a 6-ns pulsed laser system delivering 200 mJ per pulse at a repetition rate of 50 Hz. A single pulse is used to record an individual hogel in the photorefractive polymer. The writing time is reduced by several orders of magnitude, recording a 4 inch \times 4 inch sample in about 2 s, with 1-mm hogel resolution. The overall holographic recording set-up is also insensitive to vibration owing to the short pulse duration. As a result, the system is capable of operation in an industrial environment without any special need for vibration, noise or temperature control.

The holographic recording material used here is a dynamic photorefractive polymer with the following structure and composition: a copolymer with a polyacrylic backbone was used to attach pendant groups, tetraphenylidiaminobiphenyl (TPD) and carbaldehyde aniline (CAAN) in the ratio 10:1 (TPD/CAAN)²². This whole copolymer structure is then referred to as PATPD/CAAN. A plasticizer, 9-ethyl carbazole (ECZ), was added to the composite to lower the glass transition temperature and facilitate chromophore alignment in the photorefractive space charge field. The index modulation properties are enhanced by adding a fluorinated dicyanostyrene (FDCST) chromophore. In order to sensitize the photorefractive material to the nanosecond pulse, we doped the polymer with a fullerene derivative, PCBM ([6,6]-phenyl-C₆₁-butyric acid methyl ester).

The photorefractive device was made by melting the final composite PATPD/CAAN:FDCST:ECZ:PCBM (49.5:30:20:0.5 wt%) between two ITO (indium tin oxide)-coated glass plates, with 100 μ m spacers used to fix the film thickness. The ITO electrodes are used to apply an external electric field to the sample for the photorefractive effect to happen. Long-term stability against both crystallization and dielectric breakdown has been observed with this material, with samples used for months without any sign of degradation.

Figure 1 shows the dynamics of the diffracted signal from the photorefractive polymer in a four wave mixing configuration for a single 6-ns writing pulse. A single pulse is emitted at time $t = 0$ and saturates the detector (initial vertical line on the figure). The diffraction grating grows for the first 50 s, followed by a slow decay in the dark during

which the diffraction can clearly be observed for several minutes. In the actual display set-up, slow dark decay is prevented and the decay time becomes much shorter because (1) the decay time of photorefractive polymeric materials is much faster under illumination than in the dark, and (2) rewriting a new image erases the previous image. The slow (50 s) rise time dynamic, occurring on timescales much longer than the pulse duration, is due to the photorefractive process that involves several phenomena taking place after the pulse passing, such as photoconduction, charge trapping, and molecular orientation²³. We note that our device is self-developing and there is no need for additional material processing, as opposed to the silver halides or photopolymers used in static holography.

The 3D image is viewed using an incoherent coloured LED and is clearly visible under ambient room light. Pictures of a hologram recorded with the pulsed system are presented in Fig. 2. Supplementary Movie 1 shows the recording process taking place in about 2 s and camera movement further demonstrates the 3D nature of the image. The hologram fades away after a couple of minutes by natural dark decay, or it can be erased by recording the new 3D images. When recording a new image, the new interference pattern erases the old pattern, resulting in the formation of a new diffraction structure.

To achieve a multi-colour holographic 3D display, we used angular multiplexing. Up to three different holograms are written in the material at different angles and read out with different coloured LEDs (red-green-blue for full colour). The reference beam angles were set more than 10° apart to avoid crosstalk between the colours during readout. In order to precisely superimpose the different coloured diffracted beams, we used Bragg's law to determine the required 5° angle between the different object beams, which helps to compensate for the angular dispersion due to writing and reading at different wavelengths. The three holograms are recorded simultaneously, so the recording speed for the coloured holograms is the same as for monochromatic holograms.

We used orthogonal polarizations and symmetrical incidence to avoid ghost diffraction resulting from the interference between the different object beams. When beams are orthogonally polarized, they do not interfere to create an intensity modulation pattern and do not form a hologram in the material. A unique property of photorefractive polymers is that beams with incidence angles that are symmetric with

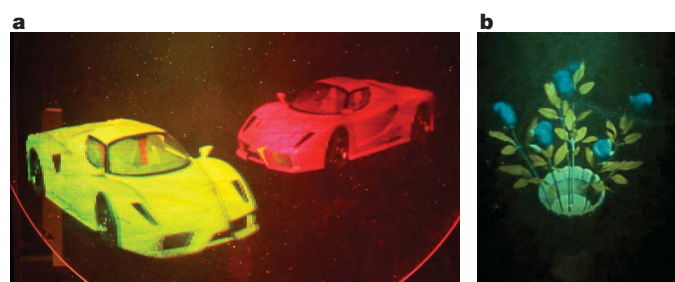


Figure 3 | Pictures of coloured holograms. a, Hologram of two model cars recorded on a 12-inch-diameter photorefractive device in HPO geometry. b, Hologram of a vase and flowers.

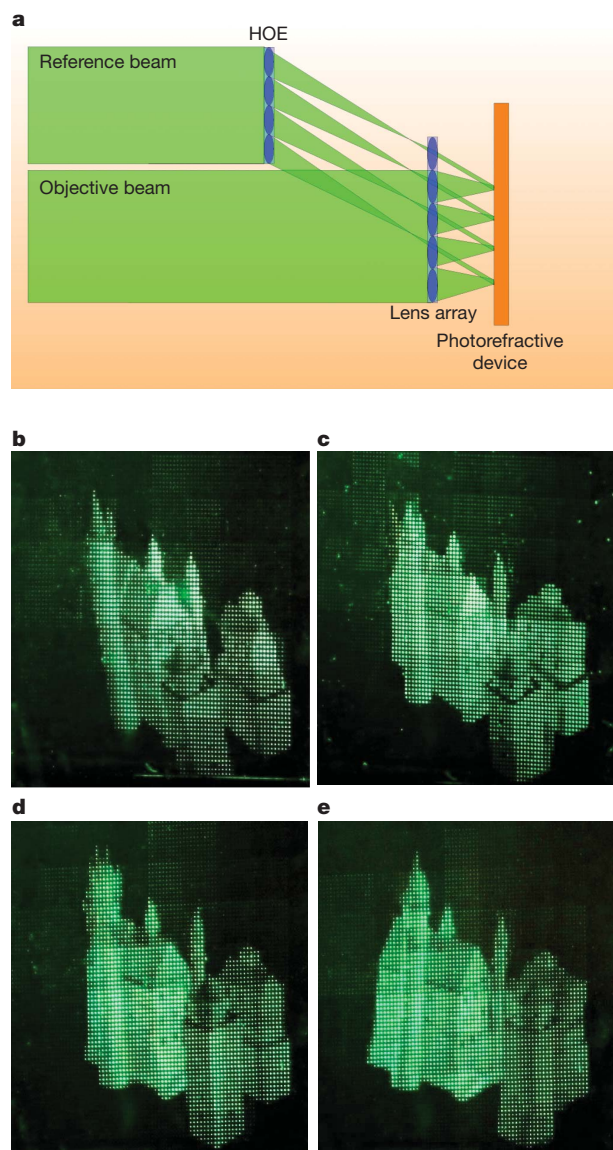


Figure 4 | Full parallax recording sketch. **a**, The lens array focuses the object beam onto the photorefractive device. The reference beam is redirected and focused by the HOE so that the lens array collimates the reference beams and they intersect the object beams in the plane of the device. **b–e**, Various perspectives of a full parallax hologram representing a castle and towers. Recording was done by simultaneously writing 100 hogels; perspectives are respectively up right (**b**), up left (**c**), down right (**d**) and down left (**e**).

respect to the direction of the external field vector do not produce a hologram. The combination of both principles allows the use of three independent pairs of recording beams for the red-green-blue colours.

Our photorefractive polymers are mostly transparent in the visible region of the spectrum, allowing us to effectively reproduce full colour in the transmission geometry. We have recorded colour holograms reproducing vivid colours, as shown in Fig. 3. The colour reproduction of the hologram can be changed according to the reading light sources used. The diffraction efficiency for each hologram is 0.5%, which is large enough to be viewed under ambient light conditions.

As perspective is mostly provided to human vision through the eyes' horizontal separation, HPO is acceptable for most 3D applications. However, full parallax is required for even better rendering or for some specific applications, such as tabletop displays that can be observed from all positions. The challenge of full parallax derives from the number of hogels to be recorded, namely the square of that needed in HPO. For example, for a 4 inch \times 4 inch screen and 1-mm resolution,

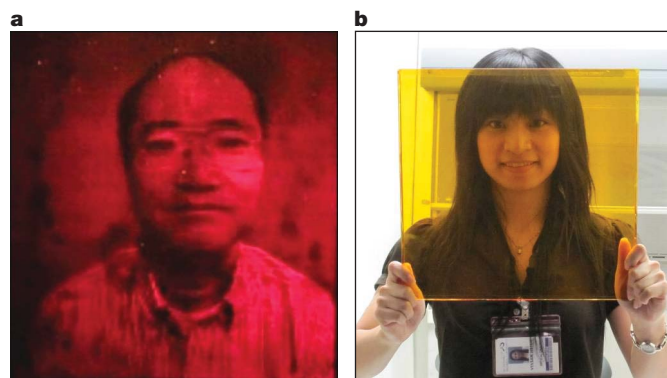


Figure 5 | Telepresence system. **a**, Picture of a hologram recorded with the 3D telepresence system. Supplementary Movie 2 shows the recording and the display of the telepresence hologram. **b**, Picture of a functional prototype of a 12 inch \times 12 inch photorefractive device, held by W.-Y.H. Images of M.Y. (**a**) and W.-Y.H. (**b**) are used with permission.

HPO requires 100 hogels whereas full parallax requires 10,000 hogels. Even with the pulsed laser system we introduce here, the recording time becomes more than 3 min for full parallax.

To overcome this limitation, we used a lens array together with a holographic optical element to spatially multiplex the recording beams. A sketch of the optical system is presented in Fig. 4a. The focal distance and the diffraction angle of the HOE were carefully designed to match the lens array focal length and lens separation so the reference beams are collimated. By using this system, we demonstrated the simultaneous recording of 100 hogels with a single pulse (Fig. 4b), decreasing the recording time by the same factor. A raster scan driven by high-speed translation stages was used to cover the entire area of the photorefractive device. The surface area illuminated with the multiplexing setup in a single exposure (1 cm²) is the same as in the HPO geometry (1 mm \times 10 cm = 1 cm²), but the spatial distribution is different (multiple squares instead of a line). Thus, the same laser source with an identical power can be used to obtain equivalent diffraction efficiency. Figure 4c shows resulting image obtained by this full parallax system.

Holographic cinematography has previously been demonstrated²⁴, but the need to process the recording media after exposure (using silver halides or photopolymer) prevents real-time operation and creates a large delay between recording and replay. With the new photorefractive polymer we describe here, quasi-real-time recording and replaying is achieved. Hogel generation in holographic stereography can be performed at video rate and does not require much computational power, as opposed to computer-generated holograms. The real-time recording and viewing, together with the lack of a requirement for extensive computational power, opens the door to new applications, including 3D telepresence—as we now demonstrate.

Our 3D telepresence set-up consists of 16 Firewire cameras that take simultaneous pictures of a real 3D scene every second. The 16 views are processed into hogel data by the host computer, and sent to the holographic recording controller through a 100 Mbit s⁻¹ Ethernet link that is used at less than 10% of its capacity. For a 4 inch \times 4 inch display, 120 hogels are processed for HPO conditions and sent every second using a general-purpose desktop PC. The pulsed holographic recording set-up is used to continuously write the hologram according to the flow of data, so that the 3D images are refreshed continuously. Once a hologram has been written, the system uses the next available hogels to update the information.

The hologram is read by light from a colour LED incident at the Bragg angle. The advantage of the holographic transmission geometry we are using is that the viewer is able to see the hologram all the time, as reading and writing occur simultaneously. The writing beams are on the opposite side of the screen with respect to the diffracted read out light; the writing beams are therefore blocked from the viewer, which is

advantageous for safety issues. Furthermore, the monochromatic writing light is prevented from reaching the viewer with an optical filter. There is no interruption in the process, with a full 3D image being continuously displayed. As described earlier, the natural (dark) decay of the hologram is longer than the decay on illumination and thus there is no noticeable difference in intensity between the previous (old) and the new 3D image. The reference beam intensity in the writing set-up is such that it erases the old image with no residual ghost image being present after refreshing. An example of a telepresence hologram is shown in Fig. 5a. Supplementary Movie 2 shows the recording and display of holographic telepresence.

We have provided here the proofs of concept for the material and the techniques of dynamic near-real-time holographic 3D display and telepresence; we expect this development to lead to new applications of holographic 3D technology. As an example, in telemedicine and especially for brain surgery, surgeons at different locations around the world could use the technique to observe in three dimensions, in real time, and to participate in the surgical procedure.

Received 4 May; accepted 14 September 2010.

1. Slinger, C., Cameron, C. & Stanley, M. Computer-generated holography as a generic display technology. *Computer* **38**, 46–53 (August 2005).
2. Ketchel, B. P. *et al.* Three-dimensional color holographic display. *Appl. Opt.* **38**, 6159–6166 (1999).
3. DeBitetto, D. J. Holographic panoramic stereograms synthesized from white light recordings. *Appl. Opt.* **8**, 1740–1741 (1969).
4. Thomas, J., Norwood, R. A. & Peyghambarian, N. Non-linear optical polymers for photorefractive applications. *J. Mater. Chem.* **19**, 7476–7489 (2009).
5. Lucente, M. *Diffraction-Specific Fringe Computation for Electro-Holography*. PhD thesis, MIT (1994).
6. Ackermann, G. K. & Eichler, J. *Holography: A Practical Approach* (Wiley-VCH, 2007).
7. Benton, S. A. & Bove, V. M. Jr. *Holographic Imaging* (Wiley-InterScience, 2008).
8. Waters, J. P. Holographic image synthesis utilizing theoretical methods. *Appl. Phys. Lett.* **9**, 405–407 (1966).
9. St., Hilaire, P., Lucente, M. & Benton, S. A. Synthetic aperture holography: a novel approach to three dimensional displays. *J. Opt. Soc. Am. A* **9**, 1969–1978 (1992).
10. Slinger, C. W. *et al.* Recent developments in computer-generated holography: toward a practical electroholography system for interactive 3d visualization. *Proc. SPIE* **5290**, 27–41 (2004).
11. Liu, Y.-Z. *et al.* High-speed full analytical holographic computations for true-life scenes. *Opt. Express* **18**, 3345–3351 (2010).
12. Häussler, R., Schwerdtner, A. & Leister, N. Large holographic displays as an alternative to stereoscopic displays. *Proc. SPIE* **6803** (2008).
13. Balogh, T. *et al.* in *Proceedings of the TMCE 2008* (eds Horváth, I. & Rus, Z.) 1–11 (Faculty of Industrial Design Engineering, Delft University of Technology, 2008).
14. Sang, X. *et al.* Demonstration of a large-size real-time full-color three-dimensional display. *Opt. Lett.* **34**, 3803–3805 (2009).
15. Paturzo, M. *et al.* Synthesis and display of dynamic holographic 3D scenes with real-world objects. *Opt. Express* **18**, 8806–8815 (2010).
16. Chen, S.-T. & Chatterjee, M. R. Implementation of a spatially multiplexed pixilated three-dimensional display by use of a holographic optical element array. *Appl. Opt.* **37**, 7504–7513 (1998).
17. Shin, S.-H. & Javidi, B. Speckle-reduced three-dimensional volume holographic display by use of integral imaging. *Appl. Opt.* **41**, 2644–2649 (2002).
18. Ketchel, B. P. *et al.* Three-dimensional color holographic display. *Appl. Opt.* **38**, 6159–6166 (1999).
19. Tay, S. *et al.* An updatable holographic three-dimensional display. *Nature* **451**, 694–698 (2008).
20. Blanche, P.-A. *et al.* An updatable holographic display for 3D visualization. *IEEE J. Display Technol.* **4** (special issue on medical displays) 424–430 (2008).
21. Peyghambarian, N., Tay, S., Blanche, P.-A., Norwood, R. & Yamamoto, M. Rewritable holographic 3D display. *Opt. Photon. News*, **19**, 22–27 (July/August 2008).
22. Thomas, J. *et al.* Bis-triaryamine polymer-based composites for photorefractive applications. *Adv. Mater.* **16**, 2032–2036 (2004).
23. Eralp, M. *et al.* Submillisecond response of a photorefractive polymer under single nanosecond pulse exposure. *Appl. Phys. Lett.* **89**, 114105 (2006).
24. DeBitetto, D. J. A front-lighted 3-D holographic movie. *Appl. Opt.* **9**, 498–499 (1970).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge support from AFOSR, DARPA and the NSF ERC Center on Integrated Access Networks (CIAN).

Author Contributions P.-A.B. and A.B. did experimental work. R.V. did modelling and software. C.C., P.W. and M.K. did experimental work. W.L., T.G., D.F., W.-Y.H., B.R., O.S. and J.T. did sample preparation. R.A.N. and Y.Y. are team leaders. N.P. did project planning and management.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.P. (nnp@u.arizona.edu).

Reversed flow of Atlantic deep water during the Last Glacial Maximum

César Negre¹, Rainer Zahn^{1,2,3}, Alexander L. Thomas⁴, Pere Masqué^{1,3}, Gideon M. Henderson⁴, Gema Martínez-Méndez^{1†}, Ian R. Hall⁵ & José L. Mas⁶

The meridional overturning circulation (MOC) of the Atlantic Ocean is considered to be one of the most important components of the climate system. This is because its warm surface currents, such as the Gulf Stream, redistribute huge amounts of energy from tropical to high latitudes and influence regional weather and climate patterns, whereas its lower limb ventilates the deep ocean and affects the storage of carbon in the abyss, away from the atmosphere. Despite its significance for future climate, the operation of the MOC under contrasting climates of the past remains controversial. Nutrient-based proxies^{1,2} and recent model simulations³ indicate that during the Last Glacial Maximum the convective activity in the North Atlantic Ocean was much weaker than at present. In contrast, rate-sensitive radiogenic $^{231}\text{Pa}/^{230}\text{Th}$ isotope ratios from the North Atlantic have been interpreted to indicate only minor changes in MOC strength^{4–6}. Here we show that the basin-scale abyssal circulation of the Atlantic Ocean was probably reversed during the Last Glacial Maximum and was dominated by northward water flow from the Southern Ocean. These conclusions are based on new high-resolution data from the South Atlantic Ocean that establish the basin-scale north to south gradient in $^{231}\text{Pa}/^{230}\text{Th}$, and thus the direction of the deep ocean circulation. Our findings are consistent with nutrient-based proxies and argue that further analysis of $^{231}\text{Pa}/^{230}\text{Th}$ outside the North Atlantic basin will enhance our understanding of past ocean circulation, provided that spatial gradients are carefully considered. This broader perspective suggests that the modern pattern of the Atlantic MOC—with a prominent southerly flow of deep waters originating in the North Atlantic—arose only during the Holocene epoch.

A characteristic feature of our present ocean circulation is the deep convection that occurs in the North Atlantic and spreads North Atlantic Deep Water (NADW) to the world's abyssal oceans. This convection is compensated by the northward flow of warm subtropical surface waters that supply the North Atlantic with large amounts of heat. Changes in MOC therefore carry profound implications for global climate. Information about the operation of the MOC before the past 100 years is obtained from palaeoceanographic proxies, such as stable carbon isotopes ($\delta^{13}\text{C}$) and trace element ratios (Cd/Ca) recorded in the biogenic carbonate of bottom-dwelling foraminifera that trace the dispersal of biologically cycled nutrients in the ocean. Mapping of these palaeo-hydrographic data suggested that under glacial conditions nutrient-poor NADW ventilated a much smaller fraction of the deep Atlantic, which was dominated by Southern Component Waters (SCW) from the Southern Ocean^{1,2,7}. This is consistent with deep-water temperature, salinity and oxygen reconstructions using independent proxy data and climate modelling^{3,8}. These proxies, however, are influenced by deep water circulation and biological nutrient cycling alike, and do not allow a quantitative reconstruction of the abyssal flow rate, which sets marine heat transport and carbon storage.

Information on the abyssal flow rate of MOC can be deduced from the radiogenic isotope pair ^{231}Pa and ^{230}Th in sea-floor sediments^{4,9,10}. This is due to their constant production in the water column from decay of dissolved uranium at a fixed $^{231}\text{Pa}/^{230}\text{Th}$ activity ratio of 0.093, and their differential solubility in the ocean (see Supplementary Information section 1 for details). Mapping of $^{231}\text{Pa}/^{230}\text{Th}$ ratios in Atlantic sea-floor sediments showed similar values in Last Glacial Maximum (LGM) and core-top sections, suggesting a southward transport of NADW during the LGM no different from, or possibly even stronger than, today⁴. Subsequent $^{231}\text{Pa}/^{230}\text{Th}$ records from the North Atlantic have indicated that measurable decreases in overturning occurred during millennial-timescale climate events in the deglaciation but indicate a still vigorous southward flow at the LGM^{5,11–13}. This is in apparent conflict with the above-mentioned reconstructions derived from nutrient-based proxies^{1,2,7}.

Interpretation of $^{231}\text{Pa}/^{230}\text{Th}$ records in the North Atlantic to date assumes that the abundance of ^{231}Pa is solely modulated by the southerly flow of NADW, which should cause $^{231}\text{Pa}/^{230}\text{Th}$ in the South Atlantic to be higher than in the North Atlantic¹⁰. Critical assessment of this interpretation has hitherto been hindered by the lack of continuous fine-scale $^{231}\text{Pa}/^{230}\text{Th}$ records from the South Atlantic. Here we present such a record, measured at multi-centennial resolution in a sediment core recovered from the Cape basin (core MD02-2594; 34° 43' S, 17° 20' E). At a water depth of 2,440 m, the site is positioned in the present-day flow path of NADW, and its southerly position makes it particularly appropriate to reconstruct the strength of NADW relative to SCW. We measured a series of complementary proxies in MD02-2594 (Fig. 1): benthic $\delta^{18}\text{O}$ for stratigraphy; benthic $\delta^{13}\text{C}$ as an indicator of the nutrient content of ambient waters and chemical ventilation; sortable silt mean grain size ($\overline{\text{SS}}$) to assess near-bottom physical flow speed¹⁴; and opal concentrations for control on ^{231}Pa scavenging by variable biogenic silica fluxes^{15,16}. Similar analyses were performed on cores further south to serve as reference for constraining $^{231}\text{Pa}/^{230}\text{Th}$ imprints of deep waters originating in the Southern Ocean (Supplementary Fig. 1).

The $^{231}\text{Pa}/^{230}\text{Th}$ ratios of core MD02-2594 display a pronounced increase from glacial values of 0.045 ± 0.005 (45–18 kyr before present, BP, $n = 15$, 1σ) to Holocene values of 0.065 ± 0.007 (10 kyr BP to present, $n = 15$). Such a shift is not observed in the associated data profiles of lithogenic sedimentation rate, authigenic U or opal flux, which argues against an imprint on the $^{231}\text{Pa}/^{230}\text{Th}$ profile by variable lithogenic flux or biological productivity. In fact, a prominent decrease is displayed by the total vertical particle flux going into the Holocene as $^{231}\text{Pa}/^{230}\text{Th}$ ratios increase. Therefore the $^{231}\text{Pa}/^{230}\text{Th}$ profile in this core primarily reflects ^{231}Pa variations driven by changes of deep Atlantic circulation. Benthic $\delta^{13}\text{C}$ and $\overline{\text{SS}}$ profiles from MD02-2594 display a clear glacial–interglacial shift, and support a major change in the basin-scale deep Atlantic circulation.

¹Institut de Ciència i Tecnologia Ambientals (ICTA), Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain. ²Institució Catalana de Recerca i Estudis Avançats (ICREA), Lluís Companys 23, 08010 Barcelona, Spain. ³Departament de Física, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain. ⁴Department of Earth Sciences, University of Oxford, South Parks Road, Oxford OX1 3AN, UK. ⁵School of Earth and Ocean Sciences, Cardiff University, Main Building, Park Place, Cardiff CF10 3YE, UK. ⁶Dipartimento Fisica Aplicada I, Escuela Universitaria Politécnica, Universidad de Sevilla, Virgen de Africa 7, 41012 Sevilla, Spain. [†]Present address: Center for Marine Environmental Sciences (MARUM), University of Bremen, Leobener Strasse, 28359 Bremen, Germany.

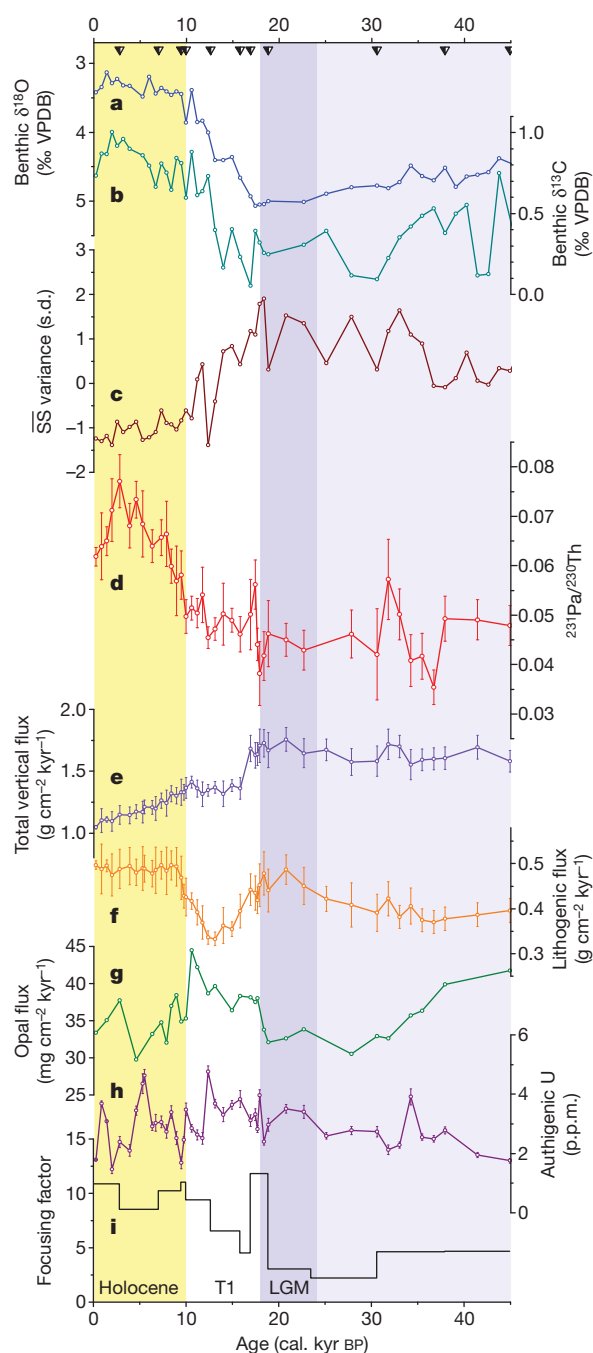


Figure 1 | Multi-proxy profiles of MD02-2594. **a, b**, Benthic stable isotope records of benthic foraminifera (*F. wuellerstorfi*) for $\delta^{18}\text{O}$ (**a**; left-hand y axis) and $\delta^{13}\text{C}$ (**b**; right-hand y axis). **c**, Sortable silt mean grain size (SS) variance. **d**, $^{231}\text{Pa}/^{230}\text{Th}$. **e**, ^{230}Th -normalized total vertical particle flux. **f**, ^{230}Th -normalized lithogenic flux. **g**, ^{230}Th -normalized opal flux. **h**, Authigenic uranium concentration. The LGM to Holocene increase of sediment focusing (**i**) leads to lower exposure of core top sediments, less dissolution of opal and therefore less masking of real opal sedimentation, ruling out an influence of increased opal fluxes on $^{231}\text{Pa}/^{230}\text{Th}$ ratios. Triangles along upper x-axis mark ^{14}C ages. Error bars give analytical s.d. Vertical shading highlights Holocene (0–10 kyr BP), T1 (10–18 kyr BP) and LGM (18–24 kyr BP).

We assess the basin-scale abyssal flow by comparing our record with existing $^{231}\text{Pa}/^{230}\text{Th}$ profiles from North Atlantic sites (Fig. 2). For this we use a composite record of OCE326-GGC5 ($33^\circ 42' \text{N}$, $57^\circ 35' \text{W}$, 4,550 m)⁵ and SU90-44 ($50^\circ 01' \text{N}$, $17^\circ 06' \text{W}$, 4,279 m)¹³, which are positioned along the present NADW path upstream from MD02-2594. Comparison of late Holocene sections of the records (past 1.5 kyr) shows a $^{231}\text{Pa}/^{230}\text{Th}$ increase from the North Atlantic

(0.055 ± 0.002 , $n = 4$) to the South Atlantic (0.064 ± 0.002 , $n = 3$), reflecting ^{231}Pa advection with the southward flow of NADW. Applying a simple radiogenic isotope bottom water flow model¹⁷ (see Methods), the late Holocene $^{231}\text{Pa}/^{230}\text{Th}$ gradient yields transit time estimates (TTEs) of 70 ± 30 yr for NADW to travel between 40°N and 35°S . These TTEs might be underestimated owing to resetting of the $^{231}\text{Pa}/^{230}\text{Th}$ signal by opal scavenging as waters pass the productive belt at the Equator, but recent work indicates that any such preferential Pa scavenging is minor in this setting during the late Holocene¹⁸. Indeed, the calculated TTE fits well with modern CFC-based observations¹⁹.

At the LGM (24–18 kyr BP) the situation is fundamentally different and the $^{231}\text{Pa}/^{230}\text{Th}$ gradient between the South and North Atlantic is reversed. Core MD02-2594 displays $^{231}\text{Pa}/^{230}\text{Th}$ ratios substantially lower than those in equivalent sections of the North Atlantic records, resulting in a meridional $^{231}\text{Pa}/^{230}\text{Th}$ gradient that increases towards the north (see Supplementary Fig. 3 for additional data from the equatorial Atlantic¹⁸). The low values at MD02-2594 also contrast with high $^{231}\text{Pa}/^{230}\text{Th}$ ratios, well above the production ratio, in sediments from the Southern Ocean that reflect ^{231}Pa enrichment due to scavenging by biogenic opal^{4,16,20}. Scavenging signals are also present in our companion record from the Agulhas Plateau that was close to, but not directly within, the northward-shifted opal belt of the LGM Southern Ocean (Supplementary Fig. 2).

The MD02-2594 $^{231}\text{Pa}/^{230}\text{Th}$ ratios hence reflect the presence of abyssal waters depleted in ^{231}Pa owing to opal scavenging²¹, which leave the Southern Ocean to ventilate the deep Atlantic. Sedimentary $^{231}\text{Pa}/^{230}\text{Th}$ ratios at the LGM that are below production ratio at North Atlantic sites, previously interpreted to reflect vigorous flow from North Atlantic sources^{4–6}, are plausibly instead due to this northward flow of low- ^{231}Pa waters. Our interpretation is consistent with indications derived from $\delta^{13}\text{C}$ and Cd/Ca that nutrient-rich bottom waters were present at deep core sites in the South and North Atlantic at the LGM^{1,2,7}. Previous $^{231}\text{Pa}/^{230}\text{Th}$ interpretations did not consider the effect of ^{231}Pa scavenging in the Southern Ocean on the ^{231}Pa imprint of southern-sourced waters^{4,5}, and therefore overestimated deep ventilation from North Atlantic sources under full-glacial conditions. The meridional $^{231}\text{Pa}/^{230}\text{Th}$ gradient of ~ 0.03 at the LGM is somewhat larger than the (reverse) gradient seen during the Holocene, and suggests a flow from the Cape Basin to the North Atlantic at about half the rate seen in the opposite direction during the Holocene (Fig. 3). A reduced MOC vigour during the LGM is consistent with recent transient atmosphere–ocean model simulations²², while increased sea-water salinity (due to enhanced sea ice formation) stimulated deep water convection in the Southern Ocean hence promoting a flow from the south and driving the LGM abyssal flow reversal³.

Along glacial Termination I (T1, 18–10 kyr BP), the South Atlantic $^{231}\text{Pa}/^{230}\text{Th}$ record displays a rather gradual increase that runs opposite to the North Atlantic profiles (Fig. 2). The only measurable step increase at our site is recorded immediately before the Heinrich 1 (H1) meltwater event in the North Atlantic. However, the coeval and more accentuated $^{231}\text{Pa}/^{230}\text{Th}$ shifts at the North Atlantic sites result in a steepened meridional gradient of 0.04 during H1. Although the $^{231}\text{Pa}/^{230}\text{Th}$ recorded in core OCE326-GGC5 during H1⁵ may be compromised by enhanced biogenic opal deposition^{23,24}, $^{231}\text{Pa}/^{230}\text{Th}$ ratios in the subpolar North Atlantic likewise reached values close to the production ratio with low opal fluxes^{12,13}. High ratios would thus reflect a weakening in the northward flow of the MOC, consistent with benthic $\delta^{13}\text{C}$, Cd/Ca and Nd isotope fingerprints in the subtropical North Atlantic that indicate the presence of waters from the south at this time^{6,25}. During the Bölling–Allerød warm period (14.5–12.8 kyr BP), $^{231}\text{Pa}/^{230}\text{Th}$ and benthic $\delta^{13}\text{C}$ in MD02-2594 close to glacial levels still suggest a deep flow dominated by ^{231}Pa -depleted SCW. However, North and South Atlantic core sites at this time were bathed by different water masses owing to a restructuring of the Atlantic MOC²⁵, and hence the meridional $^{231}\text{Pa}/^{230}\text{Th}$ gradient in this period is not diagnostic of

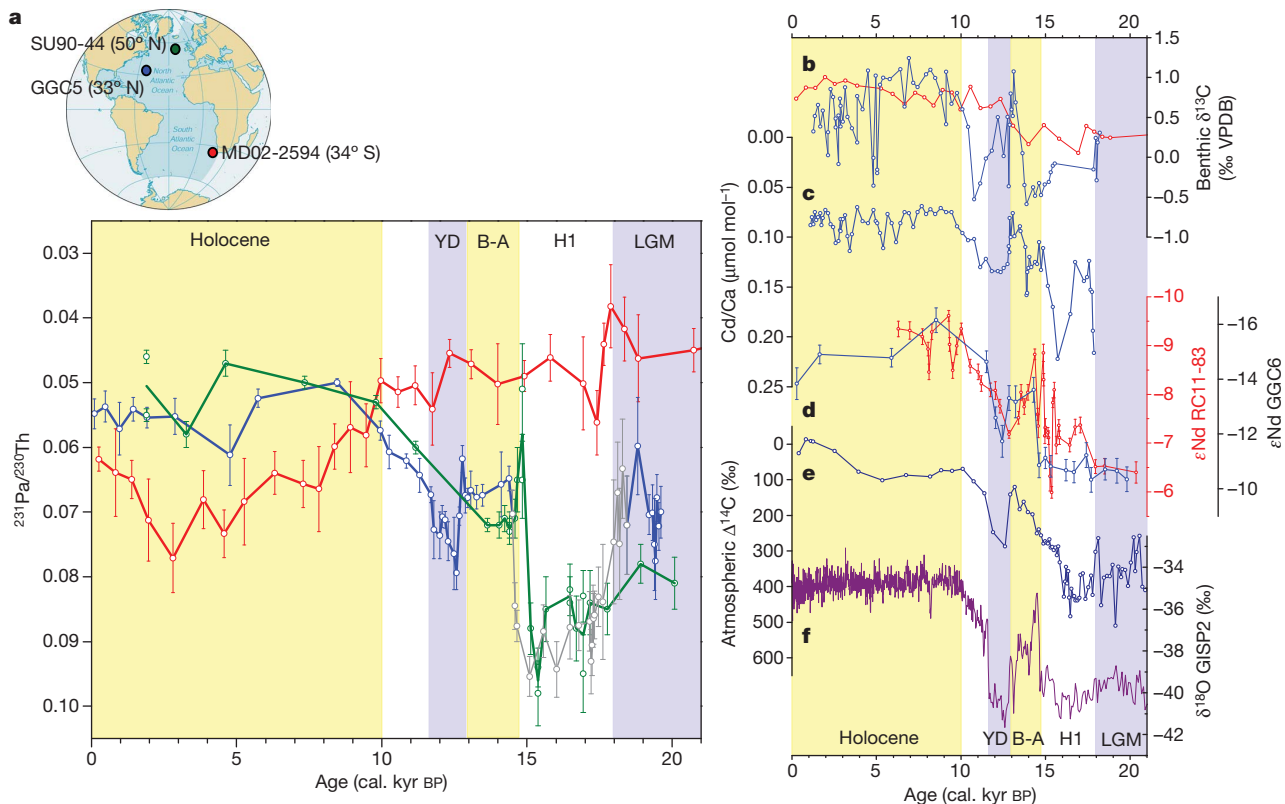


Figure 2 | MD02-2594 versus North Atlantic records. Atlantic proxy records depicting ocean circulation and climatic changes of the past 21 kyr. **a**, Top, core locations. Bottom, sedimentary $^{231}\text{Pa}/^{230}\text{Th}$ ratios from Cape Basin (MD02-2594; red; this study), eastern North Atlantic (SU90-44; green)¹³ and Bermuda Rise (OCE326-GGC5; blue)⁵. The section 18.6–14.5 kyr BP of the OCE326-GGC5 record is influenced by opal scavenging²³ and is drawn in light grey. **b**, *F. wuellerstorfi* $\delta^{13}\text{C}$ measured in Cape Basin (MD02-2594; red; this study) and Bermuda Rise (EN120-GGC1, 33° 40' N, 57° 37' W, 4,450 m; blue)²⁵

corrected for mean-ocean changes²⁹. **c**, Benthic Cd/Ca from EN120-GGC1²⁵; **d**, Nd isotope ratios (ϵ_{Nd}) in the Southeast Atlantic (RC11-83, 40° 36' S, 9° 48' E, 4,718 m; red)²⁷ and Bermuda Rise (OCE326-GGC6, 33° 41' N, 57° 35' W, 4,541 m; blue)⁶. **e**, Atmospheric $\Delta^{14}\text{C}$ profile from ODP Site 1002, Cariaco basin²⁸. **f**, $\delta^{18}\text{O}$ record from GISP2, Greenland, indicating atmospheric temperature variations³⁰. Error bars give analytical s.d. Vertical shading as Fig. 1, with Younger Dryas (YD; 11.5–12.8 kyr BP), Bölling–Allerød (B-A; 12.8–14.5 kyr BP) and H1 (~16.8 kyr BP).

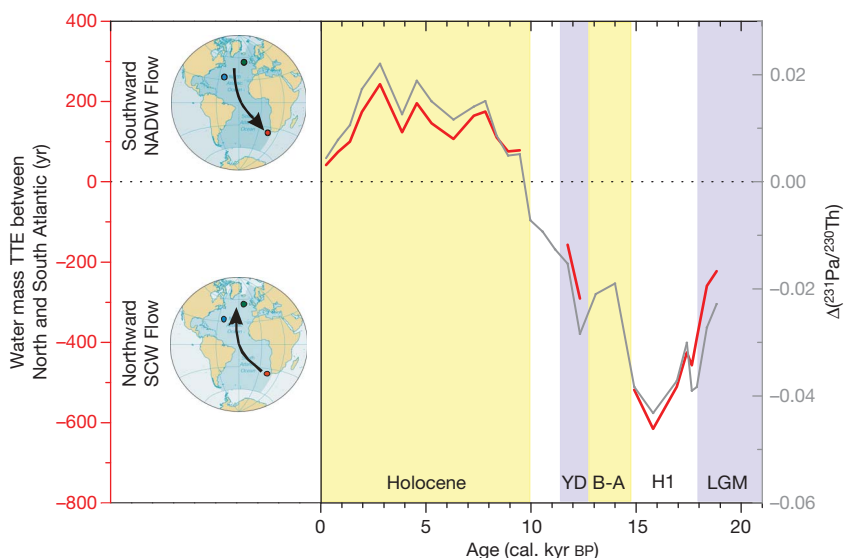


Figure 3 | Transit time estimates. Grey line (right y-axis), $^{231}\text{Pa}/^{230}\text{Th}$ gradient between MD02-2594 record and a composite of North Atlantic profiles SU90-44 and OCE326-GGC5. The composite (not shown) is an average of these two North Atlantic profiles (see Fig. 2), smoothed by a 5 point window and sampled at the time step of the MD02-2594 profile. The H1 interval in OCE326-GGC5 was excluded from the composite due to opal scavenging²³. Red lines (left y-axis), transit time estimates (TTEs) between the

mid-latitude North Atlantic and the Cape Basin in the South Atlantic, computed with an isotope bottom water flow model¹⁷. TTE was computed only for periods when North and South Atlantic sites were bathed by the same water mass. Flow patterns (shown on maps; the upper map corresponds to the Holocene epoch; the lower map to YD, H1 and LGM time periods) are reversed during the LGM, indicating abyssal water flow from the South to the North Atlantic.

the basin-scale abyssal flow vigour. A resumption of SCW is indicated by nutrient proxy records in the subtropical North Atlantic during the subsequent Younger Dryas cold event (12.8–11.5 kyr BP). The meridional $^{231}\text{Pa}/^{230}\text{Th}$ distribution here suggests that the MOC dropped back to a mode and vigour similar to those at the LGM. The absence of any larger-scale deglacial $^{231}\text{Pa}/^{230}\text{Th}$ shifts in our profile suggests that transient episodes of changing convective activity at the sources of the northern water mass were not strong enough to allow the NADW flow to overcome SCW in the South Atlantic and affect $^{231}\text{Pa}/^{230}\text{Th}$ at the MD02-2594 site, confirming numerical tracer simulations²⁶.

Within 2 kyr of the end of the Younger Dryas episode, the $^{231}\text{Pa}/^{230}\text{Th}$ ratios from the North and South Atlantic converge. At ~9.7 kyr BP, the $^{231}\text{Pa}/^{230}\text{Th}$ gradient reverses, indicating the establishment of the modern flow pattern with vigorous basin-scale southward transport of NADW that causes ^{231}Pa enrichment in the South Atlantic relative to the North Atlantic. This is consistent with other proxy data, such as benthic Cd/Ca and $\delta^{13}\text{C}$ (ref. 25) as well as Nd isotopes^{6,27}, that all indicate enhanced ventilation by a nutrient-depleted well oxygenated water mass. Near-bottom physical flow speeds also reach interglacial levels at this time, as indicated by sedimentary SS in core MD02-2594 (Fig. 1), marking the retraction of SCW as the prominent source for deep ventilation of the South Atlantic. Atmospheric ^{14}C activities reach full interglacial levels²⁸, indicating that ancient carbon was completely flushed from the ocean abyssal carbon reservoir owing to the accelerated deep ventilation from North Atlantic sources. According to the $^{231}\text{Pa}/^{230}\text{Th}$ profiles, NADW advection to the south has remained an uninterrupted feature over the whole Holocene period (10–0 kyr BP) with average TTE of 130 ± 60 yr between 40°N and 35°S .

Our South Atlantic $^{231}\text{Pa}/^{230}\text{Th}$ profile has several implications. First, our data strongly suggest that the North Atlantic $^{231}\text{Pa}/^{230}\text{Th}$ ratios at the LGM reflect the flow of abyssal waters from the Southern Ocean to the north, rather than a southward flow from North Atlantic sources as suggested before^{4,5}. Second, the LGM $^{231}\text{Pa}/^{230}\text{Th}$ gradient between our South Atlantic profile and published records from the North Atlantic is consistent with southern-sourced waters flowing northward at a rate about half the average southward-flowing NADW during the Holocene. This less vigorous deep flow is in agreement with recent numerical simulations²². Last, the absence of a $^{231}\text{Pa}/^{230}\text{Th}$ response in our profile to MOC perturbations forced by freshwater injection to the North Atlantic during the deglaciation supports tracer simulations²⁶ that demonstrate the insensitivity of our South Atlantic site to transient disruptions in the north. This therefore confirms that the progressive increase of $^{231}\text{Pa}/^{230}\text{Th}$ ratios seen in our profile from the LGM to the Holocene documents a longer-lasting reorganization of Atlantic circulation. It has previously been suggested that increased seawater salinity in the Southern Ocean⁸ in combination with surface-ocean cooling at the LGM should have stimulated enhanced convective activity at southern sources, hence potentially favouring a reversed deep abyssal flow^{3,22}. The prominent northward flow documented in the reversed $^{231}\text{Pa}/^{230}\text{Th}$ gradient at the LGM confirms these predictions and is relevant for understanding the sensitivity of the thermohaline circulation and for the calibration of climate models.

METHODS SUMMARY

For the determination of isotope abundances and $^{231}\text{Pa}/^{230}\text{Th}$, sediments were spiked and microwave-digested in a mixture of $\text{HNO}_3/\text{HCl}/\text{HF}$ and cleaned up with reverse aqua regia (Online Methods). Pa, Th and U were separated from each other using Dowex AG1-X8 resin, and measured with a Nu instruments multiple collector inductively coupled plasma mass spectrometer. The chronology of MD02-2594 was established with radiocarbon measurements of mono-specific planktonic foraminiferal samples. Analysis of $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ was performed on *Fontbotia wuellerstorfi* using a ThermoFinnigan MAT 252 mass spectrometer linked online to a single acid bath CarboKiel-II carbonate preparation device. SS measurements were undertaken on the terrigenous sub-fraction using a Coulter Multisizer III. Opal determination procedures followed extraction into Na_2CO_3 , and quantification by the colorimetric heteropoly blue method. The

content of lithogenic material was computed from ^{232}Th ; vertical rain rates of sedimentary constituents and focusing were estimated by ^{230}Th normalization.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 April; accepted 26 August 2010.

- Curry, W. B. & Oppo, D. W. Glacial water mass geometry and the distribution of $\delta^{13}\text{C}$ of ΣCO_2 in the western Atlantic Ocean. *Paleoceanography* **20**, PA1017, doi:10.1029/2004PA001021 (2005).
- Marchitto, T. M. & Broecker, W. S. Deep water mass geometry in the glacial Atlantic ocean: a review of constraints from the paleonutrient proxy Cd/Ca. *Geochim. Geophys. Res. Lett.* **7**, Q12003, doi:10.1029/2006GC001323 (2006).
- Liu, Z., Shin, S., Webb, R. S., Lewis, W. & Otto-Bliesner, B. L. Atmospheric CO_2 forcing on glacial thermohaline circulation and climate. *Geophys. Res. Lett.* **32**, L02706, doi:10.1029/2004GL021929 (2005).
- Yu, E. F., François, R. & Bacon, M. P. Similar rates of modern and last-glacial ocean thermohaline circulation inferred from radiochemical data. *Nature* **379**, 689–694 (1996).
- McManus, J. F., François, R., Gherardi, J.-M., Keigwin, L. D. & Brown-Leger, S. Collapse and rapid resumption of Atlantic meridional circulation linked to deglacial climate changes. *Nature* **428**, 834–837 (2004).
- Roberts, N. L., Piotrowski, A. M., McManus, J. F. & Keigwin, L. D. Synchronous deglacial overturning and water mass source changes. *Science* **327**, 75–78 (2010).
- Lynch-Stieglitz, J. et al. Atlantic meridional overturning circulation during the Last Glacial Maximum. *Science* **316**, 66–69 (2007).
- Adkins, J. F., McIntyre, K. & Schrag, D. P. The salinity, temperature, and $\delta^{18}\text{O}$ of the glacial deep ocean. *Science* **298**, 1769–1773 (2002).
- Henderson, G. M. & Anderson, R. F. The U-series toolbox for paleoceanography. *Rev. Mineral. Geochem.* **52**, 493–531 (2003).
- Marchal, O., François, R., Stocker, T. F. & Joos, F. Ocean thermohaline circulation and sedimentary $^{231}\text{Pa}/^{230}\text{Th}$ ratio. *Paleoceanography* **15**, 625–641 (2000).
- Gherardi, J.-M. et al. Evidence from the Northeastern Atlantic basin for variability in the rate of the meridional overturning circulation through the last deglaciation. *Earth Planet. Sci. Lett.* **240**, 710–723 (2005).
- Hall, I. R. et al. Accelerated drawdown of meridional overturning in the late-glacial Atlantic triggered by transient pre-H event freshwater perturbation. *Geophys. Res. Lett.* **33**, doi:10.1029/2006GL026239 (2006).
- Gherardi, J.-M. et al. Glacial-interglacial circulation changes inferred from $^{231}\text{Pa}/^{230}\text{Th}$ sedimentary record in the North Atlantic region. *Paleoceanography* **24**, PA2204, doi:10.1029/2008PA001696 (2009).
- McCave, I. N., Manighetti, B. & Robinson, S. G. Sortable silt and fine sediment size/composition slicing: parameters for paleocurrent speed and palaeoceanography. *Paleoceanography* **10**, 593–610 (1995).
- Chase, Z., Anderson, R. F., Fleisher, M. Q. & Kubik, P. W. The influence of particle composition and particle flux on scavenging of Th, Pa and Be in the ocean. *Earth Planet. Sci. Lett.* **204**, 215–229 (2002).
- Siddall, M. et al. $^{231}\text{Pa}/^{230}\text{Th}$ fractionation by ocean transport, biogenic particle flux and particle type. *Earth Planet. Sci. Lett.* **237**, 135–155 (2005).
- Thomas, A. L., Henderson, G. M. & McCave, I. N. Constant bottom water flow into the Indian Ocean for the past 140 ka indicated by sediment $^{231}\text{Pa}/^{230}\text{Th}$ ratios. *Paleoceanography* **22**, PA4210, doi:10.1029/2007PA001415 (2007).
- Bradtiller, L. I., Anderson, R. F., Fleisher, M. Q. & Burckle, L. H. Opal burial in the equatorial Atlantic Ocean over the last 30 ka: implications for glacial-interglacial changes in the ocean silicon cycle. *Paleoceanography* **22**, PA4216, doi:10.1029/2007PA001443 (2007).
- Huhn, O., Roether, W. & Steinfeldt, R. Age spectra in North Atlantic Deep Water along the South American continental slope, 10°N – 30°S , based on tracer observations. *Deep Sea Res.* **55**, 1252–1276 (2008).
- Walter, H.-J., Rutgers van der Loeff, M. M. & Hoelzel, H. Enhanced scavenging of ^{231}Pa relative to ^{230}Th in the South Atlantic south of the Polar Front: Implications for the use of the $^{231}\text{Pa}/^{230}\text{Th}$ ratio as a paleoproductivity proxy. *Earth Planet. Sci. Lett.* **149**, 85–100 (1997).
- Thomas, A. L., Henderson, G. M. & Robinson, L. F. Interpretation of the $^{231}\text{Pa}/^{230}\text{Th}$ paleo circulation proxy: new water-column measurements from the southwest Indian Ocean. *Earth Planet. Sci. Lett.* **241**, 493–504 (2006).
- Liu, Z. et al. Transient simulation of Last Deglaciation with a new mechanism for Bølling-Allerød warming. *Science* **325**, 310–314 (2009).
- Keigwin, L. D. & Boyle, E. A. Did North Atlantic overturning halt 17,000 years ago? *Paleoceanography* **23**, PA1101, doi:10.1029/2007PA001500 (2008).
- Lippold, J. et al. Does sedimentary $^{231}\text{Pa}/^{230}\text{Th}$ from the Bermuda Rise monitor past Atlantic meridional overturning circulation? *Geophys. Res. Lett.* **36**, L12601, doi:10.1029/2009GL038068 (2009).
- Boyle, E. A. & Keigwin, L. D. North Atlantic thermohaline circulation during the past 20,000 years linked to high-latitude surface temperature. *Nature* **330**, 35–40 (1987).
- Siddall, M. et al. Modeling the relationship between $^{231}\text{Pa}/^{230}\text{Th}$ distribution in North Atlantic sediment and Atlantic meridional overturning circulation. *Paleoceanography* **22**, PA2214, doi:10.1029/2006PA001358 (2007).
- Piotrowski, A. M., Goldstein, S. L., Hemming, S. R. & Fairbanks, R. G. Temporal relationships of carbon cycling and ocean circulation. *Science* **307**, 1933–1938 (2005).
- Hughen, K. A. et al. ^{14}C activity and global carbon cycle changes over the past 50,000 years. *Science* **303**, 202–207 (2004).
- Zahn, R. & Stuber, A. Suborbital intermediate water variability inferred from paired benthic foraminiferal Cd/Ca and $\delta^{13}\text{C}$ in the tropical West Atlantic and linking with North Atlantic climates. *Earth Planet. Sci. Lett.* **200**, 191–205 (2002).

30. Grootes, P. M. Comparison of the oxygen isotope records from GISP2 and GRIP Greenland ice cores. *Nature* **366**, 552–554 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements MD02-2594 and MD02-2588 sediment cores were provided by the International Marine Past Global Changes Study (IMAGES) and the Institut Polaire Français Paul Emile Victor (IPEV). TN057-21 and PS2489-2 samples were supplied by S. Barker and A. Martínez-García. Financial support is acknowledged from the Ministerio de Ciencia e Innovación, Spain, through scholarship AP-2004-4278 to C.N., REN2002-01958 to G.M.-M., and grant CGL2007-61579/CLI and funds from the Comer Abrupt Climate Change Foundation to R.Z. P.M. acknowledges an ICREA Academia award by the Generalitat de Catalunya.

Author Contributions R.Z. and P.M. designed the study and supervised C.N. during his PhD.; R.Z. and I.R.H. participated in the retrieval of the sediment cores; C.N. and G.M.-M. sampled the cores; C.N. processed the samples for $^{231}\text{Pa}/^{230}\text{Th}$ with help from A.L.T., J.L.M., P.M. and G.M.H.; A.L.T., G.M.H. and C.N. performed the Pa/Th/U measurements and data processing; G.M.-M. performed foraminiferal $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ analyses; I.R.H. provided SS data; C.N. analysed opal concentrations; C.N. and R.Z. wrote the paper. All authors contributed to the interpretation of the results and provided input to the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.N. (cesar@negre.us).

METHODS

Radiogenic isotope analysis. Determination of Pa, Th and U concentrations in sediments followed a recently published protocol³¹. Sediment (0.2 g) was spiked with ²³⁶U, ²²⁹Th and ²³³Pa (milked from ²³⁷Np; ref. 32), and microwave digested with a mixture of concentrated HNO₃/HCl/HF (10:4:6 ml). Solutions were dried and 9 ml of reverse aqua regia (that is, HNO₃ and HCl mixed in a molar ratio of 3:1) added as an additional cleanup step. Fluorides were removed in three evaporation-dilution steps with HNO₃, and samples were finally dissolved in 4 ml HNO₃ 7.5 M. Pa, Th and U were separated using Dowex AG1-X8 resin³³, pre-washed with HCl and Milli-Q water and preconditioned with 7.5 M HNO₃. Samples were loaded onto resin, washed with additional 4 ml of 7.5 M HNO₃, and the resin then converted to chloride form with 1.5 ml of 6 M HCl. Th, Pa and U were eluted with 6 ml of 6 M HCl, 6 M HCl + 0.05 M HF, and Milli-Q water, respectively; the Pa fraction was further purified with a repeat anion-exchange separation. Isotope abundances (²³⁰Th, ²³¹Pa, ²³²Th, ²³⁸U) were measured on a Nu Instruments MC-ICPMS at Oxford University following standard protocols²¹. In-run uncertainties (1σ) of single measurements were <2% for all isotopes. For Th and U, around 90% of the uncertainties originated from the calibration of the ²²⁹Th and ²³⁶U tracers used in this study. Reproducibility (including possible sample heterogeneity) is 7.0% for ²³¹Pa, 1.6% for ²³⁰Th, 4.6% for ²³²Th and 2.7% for ²³⁸U.

Excess sedimentary ²³¹Pa and ²³⁰Th activities were calculated correcting total concentrations for detrital and authigenic components (details in Supplementary Information section 4). Terrigenous material concentrations were inferred from ²³²Th, assuming a ²³²Th concentration of 10.7 p.p.m. in lithogenic material³⁴. Vertical rain rates of sedimentary constituents and sediment focusing were estimated by ²³⁰Th normalization³⁵.

Chronology. The chronology of core MD02-2594 is based on 11 accelerator mass spectrometry (AMS) radiocarbon measurements³⁶. The analyses were performed on mono-specific planktonic foraminiferal samples (*Globorotalia inflata*) containing more than 3 mg of carbonate. Sample preparation and ¹⁴C measurements were carried out at the National Ocean Sciences Accelerator Mass Spectrometry Facility (NOSAMS) at Woods Hole Oceanographic Institution. Radiocarbon ages were corrected applying a reservoir age of 615 ± 52 yr (ref. 37) and converted to calendar years (ref. 38).

The one-dimensional ²³¹Pa/²³⁰Th model. Deep water TTEs were computed by importing sedimentary ²³¹Pa/²³⁰Th ratios into a 1D model¹⁷. The model is a simplified representation of the processes that influence ²³¹Pa and ²³⁰Th scavenging during water mass transit. The ocean is divided into 500-m-deep boxes down to 4,000 m water depth. Production of particles settling through the underlying boxes is determined in the surface box with typical open-ocean carbonate particle flux of 0.06 g m⁻² d⁻¹ (ref. 39); distribution coefficients for Pa and Th between particulate and dissolved fractions (*K_d*) are taken from refs 15 and 16.

Stable isotopes (δ¹⁸O and δ¹³C). Benthic δ¹⁸O and δ¹³C analyses followed standard protocols⁴⁰. Sediment samples were freeze-dried to facilitate desegregation and to minimize physical damage on microfossils during wet sieving. Samples were sieved over a 63-μm screen to separate sediment coarse and fine fractions. The fine fraction (<63 μm) was oven-dried at 50 °C, weighed and used for $\overline{\text{SS}}$ analyses (see below). The coarse fraction was used for foraminiferal separation. Between 3 and 7 specimens of the epibenthic foraminifera *F. wuellerstorfi* were picked from the size fraction 250–315 μm at 10-cm steps. Cleaning procedures before stable isotope analysis involved light mechanical crushing under methanol followed by ultrasonication for 10–20 s to remove sediment coatings and release possible sediment infill. Stable isotopes were measured with a ThermoFinnigan MAT 252 mass spectrometer linked online to a single acid bath CarboKiel-II carbonate preparation device at Cardiff University. External reproducibility was monitored with an internal laboratory standard (Solenhofen Limestone) and was 0.07‰ for δ¹⁸O and 0.03‰ for δ¹³C. All isotope values are referred to the Vienna Pee Dee Belemnite scale (VPDB) through calibration to the NBS-19 carbonate standard. Benthic δ¹⁸O values are presented on the *Uvigerina* scale by adding 0.64‰ to each *F. wuellerstorfi* measurement to accommodate isotope offsets of this species from oxygen isotope equilibrium with ambient sea water⁴¹. Benthic δ¹³C values were corrected for mean-ocean changes²⁹.

Sortable silt analysis. Before $\overline{\text{SS}}$ analysis, carbonate and biogenic opal were removed from the fine fraction by dissolution in 1 M acetic acid (48 h at room temperature) and digestion in 2 M Na₂CO₃ (85 °C for 5 h), respectively. The $\overline{\text{SS}}$ grain size measurements were undertaken on the residual terrigenous sub-fraction using a Coulter Multisizer III⁴². The analytical precision ranges between 1 and 4%.

Biogenic opal analysis. Opal digestion was carried out following ref. 43. 5 ml of 10% H₂O₂ was added to 50 mg of sample in order to break down organic matter; after 30 min, an additional 5 ml of 1 M HCl were added to dissolve carbonates. Samples were sonicated for 30 min and after another 30 min, 20 ml of Milli-Q water was added and containers centrifuged at 4,500 r.p.m. for 6 min. Containers were then decanted to discard the supernatant and placed in an oven overnight at 60 °C to remove moisture. 40 ml of 2 M Na₂CO₃ were added to each sample, which were shaken and sonified, and placed in a constant-temperature bath at 85 °C to dissolve silica. After 2 and 4 h containers were shaken and placed in the bath, and after 5 h centrifuged for 6 min at 4,500 r.p.m. The supernatant was stored for analysis.

Opal analysis was performed using the colorimetric heteropoly blue method⁴⁴. 9.5 ml of Milli-Q water were added to clean polypropylene tubes, together with 0.2 ml of molybdate reagent. After 10 min, 0.2 ml of citric acid and 0.2 ml of amino-naphthol sulphonic acid were added and left for another 10 min. Next, 0.1 ml of sample solution was added to each tube and left to react for one hour. Solutions were then transferred to spectrophotometer cells and measured with a Hach Lange DR2800 spectrophotometer. A blank bracketed every sample and a standard solution was measured to monitor machine drift. Method reproducibility was ~10%.

- Negre, C. *et al.* Separation and measurement of Pa, Th, and U isotopes in marine sediments by microwave-assisted digestion and multiple collector inductively coupled plasma mass spectrometry. *Anal. Chem.* **81**, 1914–1919 (2009).
- Regelous, M., Turner, S. P., Elliott, T. R., Rostami, K. & Hawkesworth, C. J. Measurement of femtogram quantities of protactinium in silicate rock samples by multicollector inductively coupled plasma mass spectrometry. *Anal. Chem.* **76**, 3584–3589 (2004).
- Edwards, R. L., Chen, J. H. & Wasserburg, G. J. ²³⁸U–²³⁴U–²³⁰Th–²³²Th systematics and the precise measurement of time over the past 500,000 years. *Earth Planet. Sci. Lett.* **81**, 175–192 (1986).
- Taylor, S. R. & McLennan, S. M. *The Continental Crust: its Composition and Evolution* (Blackwell, 1985).
- François, R., Frank, M., Rutgers van der Loeff, M. M. & Bacon, M. P. ²³⁰Th normalisation: an essential tool for interpreting sedimentary fluxes during the late Quaternary. *Paleoceanography* **19**, PA1018, doi:10.1029/2003PA000939 (2004).
- Martínez-Méndez, G. *et al.* Contrasting multi-proxy reconstructions of surface ocean hydrography in the Agulhas Corridor and implications for the Agulhas Leakage during the last 345,000 years. *Paleoceanography* doi:10.1029/2009PA001879 (in the press).
- Southon, J., Kashgarian, M., Fortugne, M., Metivier, B. & Yim, W. E.-S. Marine reservoir corrections for the Indian Ocean and Southeast Asia. *Radiocarbon* **44**, 167–180 (2002).
- Fairbanks, R. G. *et al.* Radiocarbon calibration curve spanning 0 to 50,000 years BP based on paired ²³⁰Th/²³⁴U/²³⁸U and ¹⁴C dates on pristine corals. *Quat. Sci. Rev.* **24**, 1781–1796 (2005).
- Honjo, S. & Manganini, S. J. Annual biogenic particle fluxes to the interior of the North Atlantic Ocean: studied at 34°N 21°W and 48°N 21°W. *Deep Sea Res. II* **40**, 587–607 (1993).
- Martínez-Méndez, G., Zahn, R., Hall, I. R., Pena, L. D. & Cacho, I. 345,000-year long multi-proxy records off South Africa document variable contributions of northern versus southern component water to the deep South Atlantic. *Earth Planet. Sci. Lett.* **267**, 309–321 (2008).
- Shackleton, N. J. Attainment of isotopic equilibrium between ocean water and the benthonic foraminifera genus *Uvigerina*: isotopic changes in the ocean during the last glacial. *Colloq. Int. CNRS* **219**, 203–209 (1974).
- Bianchi, G. G., Hall, I. R., McCave, I. N. & Joseph, L. Measurements of the sortable silt current speed proxy using the Sedigraph 5100 and Coulter Multisizer II: precision and accuracy. *Sedimentology* **46**, 1001–1014 (1999).
- Mortlock, R. A. & Froelich, P. N. A simple method for the rapid determination of biogenic opal in pelagic marine sediments. *Deep-Sea Res.* **36**, 1415–1426 (1989).
- Koroleff, F. in *Methods of Seawater Analysis* 2nd edn (eds Grasshoff, K., Kremling, K. & Ehrhardt, M.) 174–183 (Verlag Chemie, Weinheim, 1983).

Higher rates of sex evolve in spatially heterogeneous environments

Lutz Becks^{1,2} & Aneil F. Agrawal¹

The evolution and maintenance of sexual reproduction has puzzled biologists for decades^{1,2}. Although this field is rich in hypotheses^{3–5}, experimental evidence is scarce. Some important experiments have demonstrated differences in evolutionary rates between sexual and asexual populations^{6–8}; other experiments have documented evolutionary changes in phenomena related to genetic mixing, such as recombination^{9,10} and selfing¹¹. However, direct experiments of the evolution of sex within populations are extremely rare (but see ref. 12). Here we use the rotifer, *Brachionus calyciflorus*, which is capable of both sexual and asexual reproduction, to test recent theory^{13–15} predicting that there is more opportunity for sex to evolve in spatially heterogeneous environments. Replicated experimental populations of rotifers were maintained in homogeneous environments, composed of either high- or low-quality food habitats, or in heterogeneous environments that consisted of a mix of the two habitats. For populations maintained in either type of homogeneous environment, the rate of sex evolves rapidly towards zero. In contrast, higher rates of sex evolve in populations experiencing spatially heterogeneous environments. The data indicate that the higher level of sex observed under heterogeneity is not due to sex being less costly or selection against sex being less efficient; rather sex is sufficiently advantageous in heterogeneous environments to overwhelm its inherent costs². Counter to some alternative theories^{16,17} for the evolution of sex, there is no evidence that genetic drift plays any part in the evolution of sex in these populations.

Sex shuffles genotypes, changing genetic associations through recombination and segregation. Sex is thought to evolve as a byproduct of the selection on these altered genetic associations⁴. All theories for the evolution of sex invoke some mechanism that maintains genetic variation because shuffling without variation does not yield any change. Much of modern theory has focused on deleterious mutations^{17–19} or host–parasite coevolution^{20–22} as the key source of genetic variation. However, a more classic explanation for the maintenance of genetic variation is spatial heterogeneity in selection^{23,24}. Several recent theoretical studies have shown that sex evolves more easily when there is spatial heterogeneity in selection^{13–15}.

If selection is the dominant evolutionary force shaping gene associations, then sex is usually disadvantageous. This is because selection leads to an excess of good allele combinations and sex destroys these combinations through recombination and segregation^{3,25}. However, in spatially heterogeneous habitats, migration, not just selection, is important in determining gene associations. Maladaptive gene combinations are constantly introduced to local populations through migration. Sex is then potentially beneficial because it helps to break down maladaptive gene associations generated by differential selection and migration^{13–15}. For sex to be favoured, the theory makes additional requirements about the nature of gene interactions (for example, locally adapted alleles should be dominant) but there are some reasons to expect these may be met¹⁴. Nonetheless, the theory is clear in predicting that the opportunity for sex to be advantageous is greater in spatially heterogeneous habitats than in homogeneous ones.

We tested this prediction in experimental populations of the monogonont rotifer *Brachionus calyciflorus* evolving in either homogeneous or heterogeneous environments. Monogonont rotifers are cyclic diploid parthenogens (that is, they normally reproduce asexually) and mixis (sexual reproduction) is induced by high rotifer densities via quorum sensing²⁶ (Supplementary Fig. 1). A preliminary study of our source population revealed ample genetic variation for the propensity to reproduce sexually, thus providing the necessary substrate for the evolution of sex (Supplementary Fig. 2).

In our experiments, rotifers were maintained in semi-continuous cultures at large population size ($N \approx 10,000$). We used two different food conditions to establish differentially selective environments (Methods). Each replicate population consisted of two subpopulations. In the homogeneous treatments, both subpopulations were of the same environmental type, that is, either both high-quality food or both low-quality food. In the heterogeneous treatment, each population was composed of one high-quality food subpopulation and one low-quality food subpopulation. Migration between subpopulations was performed by weekly manual transfer of individuals, for all treatments. Two migration rates were used corresponding to $m \approx 10\%$ and $m \approx 1\%$ per generation, assuming a generation time of about one day. Observed population densities (females, males, eggs and resting eggs) were similar across all food-quality environments (Supplementary Figs 3 and 4).

To confirm that our high-quality and low-quality food environments imposed different selection regimes, we measured fitness (lifetime per capita number of offspring) after 15 weeks of evolution of rotifers from the two homogeneous treatments (Fig. 1). We found that fitness was higher when measured in their evolved habitat than when fitness was measured in the alternative habitat, confirming differential adaptation.

We used two different methods to detect changes in the rate of sex in our experiment. First, we measured the fraction of isolated individual clones (84 per population with 42 from each subpopulation) that switched to sexual reproduction when exposed to a sex-inducing stimulus under standardized conditions (Methods). We performed this assay at the start of the experiment, after 6 weeks (about 40–45 generations) and again after 12 weeks (about 80–90 generations). The measured propensity for sex was significantly higher when migration took place in a heterogeneous environment (between high- and low-quality food conditions) than in a homogeneous environment (Fig. 2 and Supplementary Fig. 5) after 6 and 12 weeks (generalized linear model (GLM), $P < 0.001$) with a rapidly decreasing rate of sex in the spatially homogeneous environments. No significant differences were observed for the two migration rates or between the two homogeneous environments (Fig. 2).

To confirm that the differences between treatments observed under standardized assay conditions reflect real differences *in situ*, we used estimates of the percentage of sexually derived offspring (resting eggs) of total offspring (resting and amictic eggs) as a second measure of changes in the rate of sex over time and among treatments (Fig. 3). In the latter part of the experiment (from day 74 to day 109; Fig. 3b), the

¹Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, Ontario M5S 3B2, Canada. ²Department of General Ecology, Zoological Institute, Center for Biological Sciences, University of Cologne, D-50931 Köln, Germany.

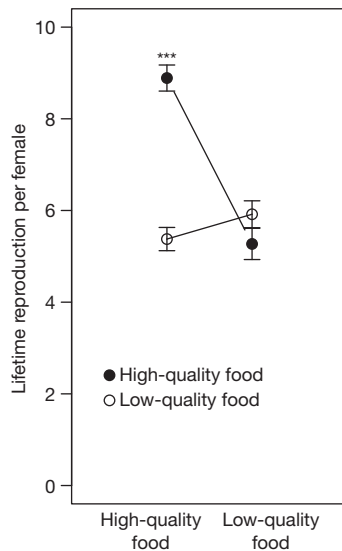


Figure 1 | Fitness in alternative environments. Lifetime reproduction (fitness) of amictic (asexual) *Brachionus calyciflorus* females was measured on individuals from populations that had evolved for 15 weeks under conditions of either high or low food quality (see text and Methods). The fitness of the same genotype of ten individual clones (third generation after isolation) was measured under high-quality food conditions (left) and low-food-quality conditions (right); $n = 18$ populations; eight populations with a migration rate of $m \approx 10\%$ and ten populations with $m \approx 1\%$ per generation. The graph shows the means \pm one standard error; GLMM $***P < 0.001$ for the high-quality food environment; the difference is not significant in the low-quality food environment. See Methods for fitness assay description.

populations reached stable densities with no significant variation in density among treatments. This permits a reasonable comparison of rates of sex among treatments during this period. The percentage of sexually derived offspring was significantly greater in the heterogeneous treatment than in the homogeneous treatments (generalized linear mixed model (GLMM): for the comparison between the heterogeneous and homogeneous high-quality food condition, $\chi^2 = 31.458$, degrees of freedom, d.f. = 1, $P < 0.001$; for the comparison between heterogeneous and homogeneous low-quality food conditions, $\chi^2 = 24.947$, d.f. = 1, $P < 0.001$). In the heterogeneous treatment, about 15% of eggs were sexually derived whereas in both of the homogeneous treatments only about 7% of eggs were sexually derived.

Both lines of evidence above indicate that sex evolves differently in heterogeneous versus homogeneous environments. Sex declines dramatically in the homogeneous treatment but little, if at all, in the heterogeneous treatment (Fig. 2 and Supplementary Fig. 5). However, there are several interpretations of this result. First, the putative benefits of sex under heterogeneity could be sufficiently large to balance its inherent costs, resulting in a higher equilibrium rate of sex than in the homogeneous environments. Second, benefits to sex may exist under heterogeneity but these are not sufficient to fully offset its costs. Consequently, sex declines in the heterogeneous treatment but at a slower rate. Third, net selection on sex does not differ between treatments (that is, there are no benefits due to heterogeneity). Rather, selection on sex is simply less efficient in the heterogeneous environment because there is more genetic variance in fitness, that is, Hill–Robertson effects²⁷ impede the elimination of alleles, causing higher rates of sex.

To distinguish amongst these possibilities, we restarted the experiment at week 14. We mixed all rotifer populations from the earlier experiment, combining all replicates of all three treatments (homogeneous high-quality food, homogeneous low-quality food, and heterogeneous) to create populations with an intermediate rate of sex (Fig. 2; the vertical line marks the mixing and the start of the second part of the experiment).

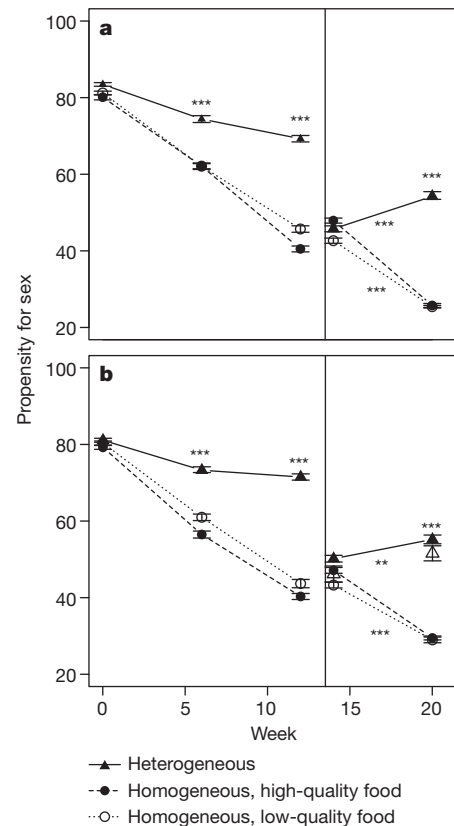


Figure 2 | Evolution of the propensity for sex in *Brachionus calyciflorus* populations from spatially heterogeneous and spatially homogeneous environments measured in a common environment. The propensity for sex was measured as the percentage of females induced into mixis (sexual reproduction) when exposed to a standardized stimulus. Vertical lines at week 14 mark the start of the second part of the experiment, when all populations were mixed and reassigned to treatments. Each data point represents the mean of 7–10 populations per treatment \pm one standard error. Migration rates are shown of $m \approx 10\%$ per generation (a) and $m \approx 1\%$ per generation (b). For both migration rates, the rate of sex is significantly greater in the heterogeneous treatment than in either homogeneous treatment in comparisons at weeks 6, 12 and 20 ($***P < 0.001$ for all comparisons). Between week 14 and 20, the rate of sex significantly increases in the heterogeneous treatment ($***P < 0.001$ for $m \approx 10\%$; $**P = 0.01$ for $m \approx 1\%$). In contrast, sex declines in the homogeneous treatments ($***P < 0.001$). Open triangles (only for $m \approx 1\%$ in b) represent heterogeneous populations evolving at ten times the standard size of $N \approx 10,000$.

Rotifers were split again into 120 populations and populations grew again for 6 weeks under the same conditions as for weeks 0–13 (food conditions, migration pattern, number of replicates) and we measured the propensity for sex again after 6 weeks of evolution (week 20 in Fig. 2). To assess the possibility of less efficient selection on sex in the heterogeneous environment due to drift-related effects, we added three additional replicates to the heterogeneous treatment (lower migration rate only) in which the population size was increased tenfold.

After an additional 6 weeks (about 40–45 generations) of evolution, the propensity for sex continued to decrease in the homogeneous populations (GLMM: in Fig. 2a, $\chi^2 = 136.98$, d.f. = 3, $P < 0.001$; in Fig. 2b, $\chi^2 = 82.903$, d.f. = 3, $P < 0.001$, for both migration rates together, $\chi^2 = 220.86$, d.f. = 2, $P < 0.001$). In contrast, the propensity for sex evolved upwards in the spatially heterogeneous populations (GLMM: in Fig. 2a, $\chi^2 = 15.41$, d.f. = 2, $P < 0.001$; in Fig. 2b, $\chi^2 = 10.72$, d.f. = 1, $P = 0.01$; for both together, $\chi^2 = 20.3$, d.f. = 1, $P < 0.001$). This result indicates that the advantages to sex outweigh its costs under spatial heterogeneity. Moreover, there were no differences between the larger and the smaller population sizes (Fig. 2b, open

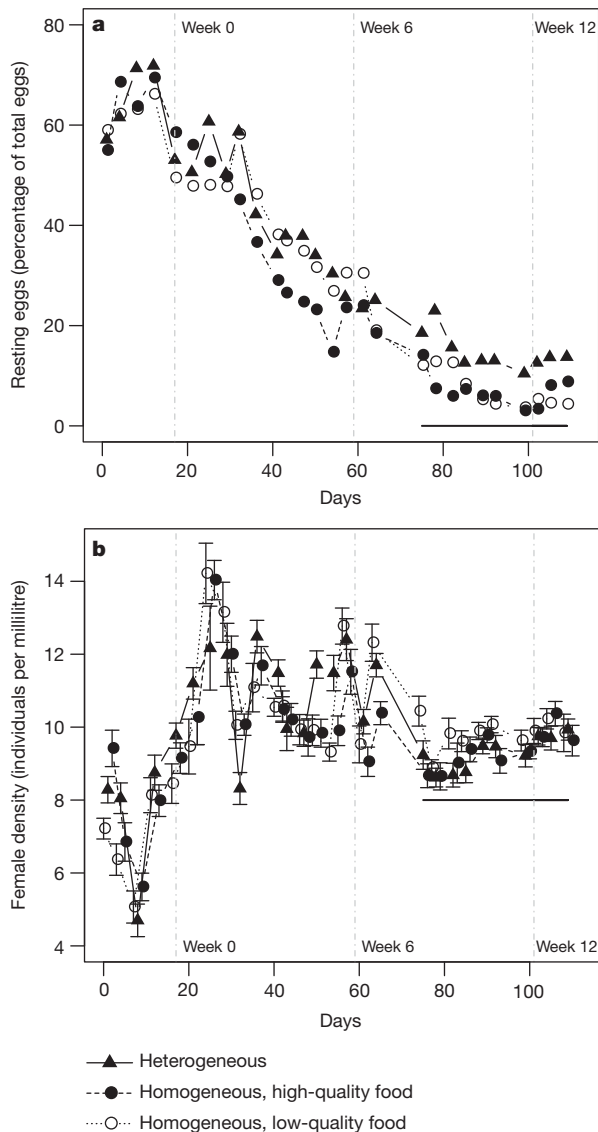


Figure 3 | In situ measure of the rate of sex in *Brachionus calyciflorus* populations from spatially heterogeneous and spatially homogeneous environments measured as the fraction of sexually derived offspring (resting eggs) of total offspring. Mean percentages of resting eggs out of all eggs in *Brachionus calyciflorus* populations (a) and mean female densities (b) are plotted over time. Error bars represent \pm one standard error. The populations appear to reach demographic equilibrium after about 75 days. After this point (days 75–109), female densities are very similar across treatments, thereby permitting a reasonable comparison for rates of sex. During this period, the percentage of sexually produced offspring was significantly higher in the heterogeneous treatment than in the homogeneous treatments (GLMM: for the heterogeneous versus homogeneous high-quality food condition, $\chi^2 = 31.458$, d.f. = 1, $P < 0.001$; for the heterogeneous versus homogeneous low-quality food condition, $\chi^2 = 24.947$, d.f. = 1, $P < 0.001$). Vertical lines mark the time points at which the propensity for sex was measured in a common environment (Fig. 2) and horizontal lines mark the data used for the comparison of resting egg fraction between treatments.

triangles), suggesting that drift-based hypotheses for the evolution of sex^{16–18} are not responsible for these results.

In this experimental system, there is the ‘twofold’ cost of meiosis²⁸ as well as a time cost because sexual reproduction takes longer than asexual reproduction. These costs tend to keep the equilibrium level of sex quite low and seem to favour complete asexuality under homogeneous conditions. Even with spatial heterogeneity, the equilibrium rate of sex is low. Nonetheless, non-zero rates of sex evolve with spatial heterogeneity despite the substantial costs. These results are therefore consistent with

recent theory predicting that there is more opportunity for sex to be favoured in spatially heterogeneous environments^{13–15}. However, this experiment was not designed to determine the mechanisms favouring higher levels of sex, so caution should be used in making interpretations in this regard. Below we consider alternative interpretations.

It is possible that higher levels of sex evolved here as a correlated response to selection for resting eggs. This could occur if resting eggs survive better than amictic eggs in transfers from one habitat type to another. If this were the case, then we would expect higher levels of sex when there was more movement between habitats because the strength of selection should be directly proportional to the probability of experiencing an environmental change, which is equivalent to the migration rate. However, there is little difference in the rate of sex between the two migration rates. The similarity between migration treatments is also somewhat surprising with respect to migration models of sex^{13–15}, though less so. Such models predict that sex is most strongly favoured at some intermediate level of migration, where the ‘optimal’ migration rate depends on the genetic architecture of the locally adapted traits. It is possible that the equilibrium level of sex may be fairly robust to quantitative differences in migration rate under the right genetic conditions.

Adaptation to new environments is also thought to facilitate the evolution of sex^{2,10,16}. Both habitats used here were slightly different to previous culture conditions, so adaptation is likely to have occurred in all treatments. Interestingly, sex seems to increase initially (days 0 to 16 in Fig. 3a) in all treatments, despite low density during this period, then later declines. This observation is consistent with sex being advantageous when the rate of adaptation is likely to be highest but then becoming less favoured (owing to its costs) after adaptation slows. Although this pattern is interesting, it seems unlikely that ‘sex for adaptation’ explains the differences that develop between homogeneous and heterogeneous treatments over the longer term.

Recent theory has shown that sex may evolve as a way for genes to escape bad genetic backgrounds^{3,29}; this is known as ‘fitness-associated’ sex. With spatial heterogeneity, maladaptive genotypes are constantly introduced through migration, providing more opportunity for fitness-associated sex to evolve. Under this hypothesis, we would expect that genotypes have more sex when tested in the alternative environment from which they were collected. However, we find no support for this prediction. Clones collected from the two different habitats in the heterogeneous environment show no difference with respect to the rate of sex when tested under the same environment (Supplementary Fig. 6).

The evolution of sex has been one of the enduring problems of evolutionary biology. Although there has been a large amount of theory, experimental tests of sex itself have been limited to comparisons of sexual and asexual populations^{6,7}. Experimental evolution has been largely neglected as a means of examining the more perplexing question of how rates of sex evolve within populations (but see ref. 12). Our experiment demonstrates that rates of sex do evolve within experimental populations and that non-zero rates of sex can be favoured even in the face of real costs. This suggests that this approach can be applied to a variety of facultatively sexual organisms (able to reproduce both sexually and asexually) to test the predictions of different theoretical models. Future work should not only identify the types of conditions that favour the evolution of sex but also examine the population genetic mechanisms by which these benefits arise⁴. By doing so, we can begin bridging the sizeable gap between theory and the empirical patterns observed in nature.

It is worth noting that the rate of sex declined in all treatments compared to the initial state (Fig. 3). Because these experiments were started with field-collected organisms, this observation suggests that the equilibrium rate of sex in nature is higher than in any of the laboratory environments used here. Whether this discrepancy is due to greater environmental heterogeneity in the field or other factors (such as parasites or more targets of selection) is unknown and presents a challenge for future studies.

METHODS SUMMARY

Creation of experimental populations. Populations were initiated from a laboratory population that had been recently hatched from a large number of field-collected resting eggs. Rotifer–algae populations were established in 500-millilitre batch cultures under two sets of conditions to create different selection regimes: high-quality food conditions and low-quality food conditions. 10% of each culture (including rotifers and algae) was replaced twice per week with the respective algae solution (2×10^6 cells ml^{-1}) taken from constant chemostat cultures. Each population was composed of two subpopulations: for the homogeneous treatment, both were either from low-quality food or both from high-quality food conditions, and for the heterogeneous treatment, the subpopulations were of opposite types. Once per week, 50% or 5% of the rotifers (females, males, eggs, resting eggs) were exchanged between the two subpopulations. Assuming a generation time of 1 to 2 days, these migration regimes are equivalent to approximately $m \approx 10\%$ and $m \approx 1\%$ per generation, respectively.

Assay of propensity to reproduce sexually. The propensity for sex was measured by isolating 42 (amictic) clones from each subpopulation (84 clones per population). Clones were individually transferred into single wells with 10 ml of high-quality food and maintained in these conditions for two amictic generations to eliminate environmental effects. For the mixis assay, two neonates of the third generation after isolation were individually transferred to a single cell of a 96-well plate with conditioned medium (sex-inducing stimulus) at two different concentrations. Conditioned medium was freshly prepared from a dense laboratory rotifer stock culture by removing rotifers from the medium and diluting it with fresh medium to a stimulus concentration representative of 22.5 females per millilitre. After the first day of the assay the initial females were removed and the offspring were scored as amictic or mictic by their produced offspring: males are produced by sexual females and females by amictic females. For the assays at the start, after 6 and 12 weeks, an additional set of tests using a lower concentration of sex-inducing stimulus (12 females per millilitre) were also performed. Results were qualitatively similar (Supplementary Fig. 5).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 June; accepted 24 August 2010.

Published online 13 October 2010.

1. Bell, G. *The Masterpiece of Nature: The Evolution and Genetics of Sexuality* (University of California Press, 1982).
2. Maynard Smith, J. *The Evolution of Sex* (Cambridge University Press, 1978).
3. Otto, S. P. The evolutionary enigma of sex. *Am. Nat.* **174**, S1–S14 (2009).
4. Agrawal, A. F. Evolution of sex: why do organisms shuffle their genotypes? *Curr. Biol.* **16**, R696–R704 (2006).
5. Kondrashov, A. S. Classification of hypotheses on the advantage of amphimixis. *J. Hered.* **84**, 372–387 (1993).
6. Colegrave, N. Sex releases the speed limit on evolution. *Nature* **420**, 664–666 (2002).
7. Goddard, M. R., Godfray, H. C. J. & Burt, A. Sex increases the efficacy of natural selection in experimental yeast populations. *Nature* **434**, 636–640 (2005).
8. Poon, A. & Chao, L. Drift increases the advantage of sex in RNA bacteriophage Phi 6. *Genetics* **166**, 19–24 (2004).
9. Korol, A. B. & Iliadi, K. G. Increased recombination frequencies resulting from directional selection for geotaxis in *Drosophila*. *Heredity* **72**, 64–68 (1994).
10. Otto, S. P. & Barton, N. H. Selection for recombination in small populations. *Evolution* **55**, 1921–1931 (2001).
11. Morran, L. T., Parmenter, M. D. & Phillips, P. C. Mutation load and rapid adaptation favour outcrossing over self-fertilization. *Nature* **462**, 350–352 (2009).
12. Wolf, H. G., Wohrmann, K. & Tomiuk, J. Experimental evidence for the adaptive value of sexual reproduction. *Genetica* **72**, 151–159 (1987).
13. Lenormand, T. & Otto, S. P. The evolution of recombination in a heterogeneous environment. *Genetics* **156**, 423–438 (2000).
14. Agrawal, A. F. Spatial heterogeneity and the evolution of sex in diploids. *Am. Nat.* **174**, S54–S70 (2009).
15. Pytkov, K. V., Zhivotovsky, L. A. & Feldman, M. W. Migration versus mutation in the evolution of recombination under multilocus selection. *Genet. Res.* **71**, 247–256 (1998).
16. Felsenstein, J. & Yokoyama, S. Evolutionary advantage of recombination. 2. Individual selection for recombination. *Genetics* **83**, 845–859 (1976).
17. Keightley, P. D. & Otto, S. P. Interference among deleterious mutations favours sex and recombination in finite populations. *Nature* **443**, 89–92 (2006).
18. Roze, D. & Michod, R. D. Deleterious mutations and selection for sex in finite diploid populations. *Genetics* **184**, 1095–1112 (2010).
19. Kondrashov, A. S. Deleterious mutations as an evolutionary factor. 1. The advantage of recombination. *Genet. Res.* **44**, 199–217 (1984).
20. Agrawal, A. F. Differences between selection on sex versus recombination in Red Queen models with diploid hosts. *Evolution* **63**, 2131–2141 (2009).
21. Otto, S. P. & Nuismer, S. L. Species interactions and the evolution of sex. *Science* **304**, 1018–1020 (2004).
22. Peters, A. D. & Lively, C. M. Short- and long-term benefits and detriments to recombination under antagonistic coevolution. *J. Evol. Biol.* **20**, 1206–1217 (2007).
23. Hedrick, P. W. Genetic polymorphism in heterogeneous environment—a decade later. *Annu. Rev. Ecol. Syst.* **17**, 535–566 (1986).
24. Felsenstein, J. Theoretical population-genetics of variable selection and migration. *Annu. Rev. Genet.* **10**, 253–280 (1976).
25. Feldman, M. W., Otto, S. P. & Christiansen, F. B. Population genetic perspectives on the evolution of recombination. *Annu. Rev. Genet.* **30**, 261–295 (1997).
26. Gilbert, J. J. Specificity of crowding response that induces sexuality in the rotifer *Brachionus*. *Limnol. Oceanogr.* **48**, 1297–1303 (2003).
27. Hill, W. G. & Robertson, A. Effects of linkage on limits to artificial selection. *Genet. Res.* **8**, 269–294 (1966).
28. Williams, G. C. *Sex and Evolution* (Princeton University Press, 1975).
29. Hadany, L. & Otto, S. P. The evolution of condition-dependent sex in the face of high costs. *Genetics* **176**, 1713–1727 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Kearns and N. G. Hairston Jr for providing the sediment with resting eggs from which the *Brachionus* cultures were established. This research was funded by the Volkswagen Foundation to L.B. (I/83 517) and the Natural Sciences and Engineering Research Council (Canada) to A.F.A.

Author Contributions L.B. and A.F.A. conceived and designed the study, L.B. performed experiments, L.B. and A.F.A. discussed and analysed the results, and shared the writing of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.B. (lutz.becks@utoronto.ca).

METHODS

All cultures were inoculated from a *Brachionus calyciflorus* laboratory stock culture, which was started in June 2009 from field-collected resting egg hatches. Rotifer–algae cultures were kept at $25 \pm 1.5^\circ\text{C}$ (12 h dark/12 h light) in tissue culture flasks (Sarsted) and moved randomly three times per week across three shelves of an incubator. Each population consisted of two subpopulations. Each subpopulation was maintained in a 500-ml batch culture at a density of about ten rotifers per millilitre, that is, about 5,000 rotifers per subpopulation giving $N \approx 10,000$ per population (detailed density data provided in Supplementary Fig. 3). Each subpopulation was one of two habitat types: high-quality food conditions or low-quality food conditions. 10% of each subpopulation (including rotifers and algae) was removed twice per week and an equivalent volume of the appropriate algae solution (below) was replaced. From the removed sample, rotifer densities (females, males, eggs, resting eggs) were enumerated under a stereoscope. For the replacement solution, algae were taken from long-term chemostat cultures to ensure constant food conditions over the course of experiment (high-quality food condition: nitrogen concentration in medium = $1,000\ \mu\text{M}$ per litre; low-quality food condition: nitrogen concentration in medium = $160\ \mu\text{M}$ with an additional $0.5\ \text{g}\ \text{ml}^{-1}$ NaCl per litre); both chemostats were inoculated from the same stock four weeks before the start of the experiment (SAG 278-3, Algae Culture Collection University of Göttingen). The replacement algal solution was prepared by diluting either of the two chemostat cultures to a concentration of 2×10^6 algal cells per millilitre in nitrogen-free inorganic medium (modified after ref. 30 with $0.5\ \text{g}\ \text{L}^{-1}$ NaCl added to the low-quality food environment).

Asexual rotifer females produced about one offspring per day and juveniles started reproducing a few hours after hatching (observed by L.B.). Migration of rotifers (females, males, eggs, resting eggs) between two subpopulations took place once per week by filtering out the rotifers (adults and eggs) from 50% or 5% of the volume from each subpopulation and exchanging the rotifers between subpopulation pairs. Assuming a generation time of 1 to 2 days, a 50% or 5% weekly exchange corresponds to $m \approx 10\%$ or $m \approx 1\%$ per generation, respectively. Migration started two weeks (day 18) after rotifer inoculation.

At week 15 (week 13 of migration), ten clones were isolated from each subpopulation and transferred individually to 10 ml of high-quality food. Two neonates of the third generation after isolation were used to assay lifetime reproduction (fitness) under either high-quality food or low-quality food conditions. The number of offspring of females was recorded daily and females were transferred to new wells with fresh medium at the same time until the females died.

Sex stimulus in standardized environment. Mixis in *Brachionus calyciflorus* is stimulated through a mixis-inducing protein that is excreted by *Brachionus* females and accumulates in the medium²⁶ (Supplementary Fig. 1). Conditioned medium (sex stimulus) was prepared freshly from a dense laboratory stock culture by removing rotifers (repeated filtration through $10\ \mu\text{m}$ mesh) from the medium and diluting it with fresh medium to a desired 'equivalent rotifer density' of 22.5 or 12 females per millilitre (ref. 26). For the second part of the experiment after week 14, rotifers were only tested at 22.5 females per millilitre.

The propensity for sex was measured by isolating 42 asexual (amictic) clones from each subpopulation (84 clones per population). Clones were individually transferred into single wells with 10 ml of high-quality food and maintained under

these conditions for two generations. For the mixis assay, two neonates of the third generation after isolation were individually transferred to a single cell of a 96-well plate with conditioned medium at the two different concentrations. After the first day of the assay the initial females were removed and her offspring were scored as amictic or mictic by the type of offspring that they produced. Amictic (asexual) females produce female offspring. Sexual females produce only haploid males because they are unmated in this assay (Supplementary Fig. 1).

Creation of populations with a low initial rate of sex. After week 13, experimental populations from all treatments were mixed. Because two-thirds of these populations were from the homogeneous treatment, the average rate of sex in this newly mixed population was quite low. New experimental populations were created from this mixed population and distributed amongst the same set of treatments as in the original experiment. To assess whether genetic drift was affecting evolution in the heterogeneous treatment, three larger populations were added to this treatment (for the lower migration rate only). The larger populations grew under the same conditions but each population consisted of 20 subpopulations (ten of each type) rather than two. For migration, 10% of the volume of each of the ten subpopulations of the same habitat type was removed and pooled. Half of this volume was equally distributed back to the same subpopulations (within-habitat migration) and the remainder was equally distributed amongst the ten alternative habitat subpopulations (between-habitat migration).

Data analysis. Multivariate analyses were performed in the R statistical environment using the lmer4 package^{31,32}. Fitness data were analysed using generalized mixed models with environmental origin (the high- or low-quality food environment to which the populations adapted) as a fixed effect and replicated population as a random effect for the comparison of the mean fitness tested within one environment (either original or novel in Fig. 1; GLMM with quasi-Poisson error structure). Differences among environments (heterogeneous, homogeneous low-quality food, homogeneous high-quality food) in the percentages of mixis-induced females in the mixis assay (Fig. 2) were tested by using a GLM (with binomial error distribution). The effect of population size (Hill–Robertson effect) on the rate of sex within the heterogeneous treatment was tested in the same way. To test for differences in the rate of sex between the two migration rates, we compared generalized linear models with and without the migration term. For each environmental treatment, GLMM were used to test the change in the rate of sex between weeks 14 and 20, with time as a fixed effect and replicated population as a random effect (for the heterogeneous treatment, both population sizes were included). Differences in the percentage of resting eggs out of all eggs (*in situ*) were estimated from resting egg and amictic egg densities (days 75 to 109); statistical differences were determined by comparing generalized mixed models with and without treatment, treating replicated population as a random effect (quasi-binomial error correction).

30. Fussmann, G. F., Ellner, S. P., Shertzer, K. W. & Hairston, N. G. Jr. Crossing the Hopf bifurcation in a live predator–prey system. *Science* **290**, 1358–1360 (2000).
31. Team, R. D. C. R.: *a Language and Environment for Statistical Computing* (R Foundation for Statistical Computing Vienna, 2009).
32. Bates, D. & Maechler, M. *lme4: Linear Mixed-Effects Models Using Eigen and S4*. R package version 0.999375-31 (<http://CRAN.R-project.org/package=lme4>) (2009).

Antibodies to human serum amyloid P component eliminate visceral amyloid deposits

Karl Bodin^{1*}, Stephan Ellmerich^{1*}, Melvyn C. Kahan¹, Glenys A. Tennent¹, Andrzej Loesch¹, Janet A. Gilbertson¹, Winston L. Hutchinson¹, Palma P. Mangione^{1,2}, J. Ruth Gallimore¹, David J. Millar¹, Shane Minogue³, Amar P. Dhillon⁴, Graham W. Taylor¹, Arthur R. Bradwell^{5,6}, Aviva Petrie⁷, Julian D. Gillmore¹, Vittorio Bellotti^{1,2}, Marina Botto⁸, Philip N. Hawkins¹ & Mark B. Pepys¹

Accumulation of amyloid fibrils in the viscera and connective tissues causes systemic amyloidosis, which is responsible for about one in a thousand deaths in developed countries¹. Localized amyloid can also have serious consequences; for example, cerebral amyloid angiopathy is an important cause of haemorrhagic stroke. The clinical presentations of amyloidosis are extremely diverse and the diagnosis is rarely made before significant organ damage is present¹. There is therefore a major unmet need for therapy that safely promotes the clearance of established amyloid deposits. Over 20 different amyloid fibril proteins are responsible for different forms of clinically significant amyloidosis and treatments that substantially reduce the abundance of the respective amyloid fibril precursor proteins can arrest amyloid accumulation¹. Unfortunately, control of fibril-protein production is not possible in some forms of amyloidosis and in others it is often slow and hazardous¹. There is no therapy that directly targets amyloid deposits for enhanced clearance. However, all amyloid deposits contain the normal, non-fibrillar plasma glycoprotein, serum amyloid P component (SAP)^{2,3}. Here we show that administration of anti-human-SAP antibodies to mice with amyloid deposits containing human SAP triggers a potent, complement-dependent, macrophage-derived giant cell reaction that swiftly removes massive visceral amyloid deposits without adverse effects. Anti-SAP-antibody treatment is clinically feasible because circulating human SAP can be depleted in patients by the bis-D-proline compound CPHPC⁴, thereby enabling injected anti-SAP antibodies to reach residual SAP in the amyloid deposits. The unprecedented capacity of this novel combined therapy to eliminate amyloid deposits should be applicable to all forms of systemic and local amyloidosis.

SAP is selectively concentrated in amyloid deposits by its avid binding to all amyloid fibril types^{2,3}. SAP binding stabilizes amyloid fibrils, protects them from proteolysis *in vitro*⁵ and contributes to the pathogenesis of systemic amyloidosis *in vivo*⁶. Therefore, we developed a new bis-D-proline compound, (R)-1-[6-[(R)-2-carboxy-pyrrolidin-1-yl]-6-oxo-hexanoyl]pyrrolidine-2-carboxylic acid (CPHPC), which is bound with high affinity by human SAP and triggers its rapid clearance by the liver, thereby depleting circulating SAP by more than 90% for as long as the drug is administered^{4,7}. However, some SAP remains bound to amyloid even after months of CPHPC treatment⁷. Here we show that targeting this residual SAP with immunoglobulin-G (IgG) antibodies triggers the body's potent phagocytic clearance mechanisms (Supplementary Information, section 1).

Splenic and hepatic amyloid A protein (AA) amyloid deposition, closely resembling human systemic AA amyloidosis, was induced by chronic inflammation in C57BL/6 mice deficient in mouse SAP but transgenic for human SAP⁴. Human SAP is present in their circulation,

in normal extracellular matrix^{8,9} (Supplementary Information, section 2) and in the amyloid deposits (Supplementary Information, section 3), just as in humans. Amyloid was quantified in each mouse by whole-body retention of ¹²⁵I-SAP (ref. 10) and the mice were allocated to three groups closely matched for age, sex and amyloid load. The model closely reflects clinical amyloidosis because human SAP binds much more avidly to amyloid than does mouse SAP¹⁰, and CPHPC depletes circulating human but not mouse SAP *in vivo*⁴. Two groups of mice then received CPHPC at 1 mg ml⁻¹ in their drinking water for the rest of the experiment. Circulating human SAP was depleted but, as in humans treated with CPHPC, significant amounts of SAP remained in the amyloid deposits (Supplementary Information, section 4). Five days after starting on CPHPC, one group received a single intraperitoneal injection of 50 mg of the IgG fraction of monospecific polyclonal sheep anti-human-SAP antiserum, containing 7 mg of anti-SAP antibody. A control group received 50 mg of unrelated sheep IgG (Supplementary Information, section 4). The third group received no treatment and thus controlled for spontaneous regression of AA amyloid¹¹. Twenty-eight days after the antibody or control IgG injection, the visceral amyloid load was scored histologically and human SAP was quantified in the individual sera and organs (Supplementary Information, section 4).

There was markedly less amyloid after treatment with CPHPC plus anti-SAP antibody than in the other two groups but there was no difference between CPHPC alone and no treatment (Fig. 1 and Supplementary Information, section 4). Apart from the amyloid deposits there were no other significant histological abnormalities in any animal. Anti-SAP antibody thus produced remarkable regression of amyloid with no disruption to the normal parenchymal or connective-tissue structure of the liver, spleen or other organs. Furthermore, there were no clinical or biochemical adverse effects, no mice died during the experiment and body weights remained constant (Supplementary Information, section 5).

Systemic amyloid deposits are characteristically acellular with no surrounding inflammatory reaction (Figs 2l, p, 3a). However, by 24 h after injection of anti-SAP antibody all the deposits were densely infiltrated with mononuclear inflammatory cells and some granulocytes (Fig. 2b, m, q). Most infiltrating cells stained strongly with antibody to F4/80, a global macrophage marker (Fig. 2b). No such staining was present in the amyloid deposits of mice not receiving anti-SAP. On day 2, the macrophages surrounding the amyloid were fusing to form multinucleate giant cells and stained strongly for CD68, a marker of phagocyte endocytotic activity that co-localized with staining for the amyloid fibril AA protein and mouse complement component C3 (Figs 2e–k, 3b, c and Supplementary Information, section 6). By day 4 the deposits were less abundant and were fragmented by numerous multinucleate giant cells surrounding and engulfing islands of amyloid (Fig. 2c–e, n, r). At day 7,

¹Centre for Amyloidosis and Acute Phase Proteins, Division of Medicine, University College London, Royal Free Campus, Rowland Hill Street, London NW3 2PF, UK. ²Dipartimento di Biochimica, Università di Pavia, Via Taramelli 3b, 27100 Pavia, Italy. ³Centre for Molecular Cell Biology, Division of Medicine, University College London, Royal Free Campus, Rowland Hill Street, London NW3 2PF, UK. ⁴Department of Histopathology, University College London, Royal Free Campus, Rowland Hill Street, London NW3 2PF, UK. ⁵Department of Immunity and Infection, The Medical School, University of Birmingham, Birmingham B15 2TT, UK. ⁶The Binding Site Ltd, Birmingham B14 4ZB, UK. ⁷Biostatistics Unit, UCL Eastman Dental Institute, 256 Grays Inn Road, London WC1X 8LD, UK. ⁸Rheumatology Section, Faculty of Medicine, Imperial College London, Hammersmith Campus, Du Cane Road, London W12 0NN, UK.

*These authors contributed equally to this work.

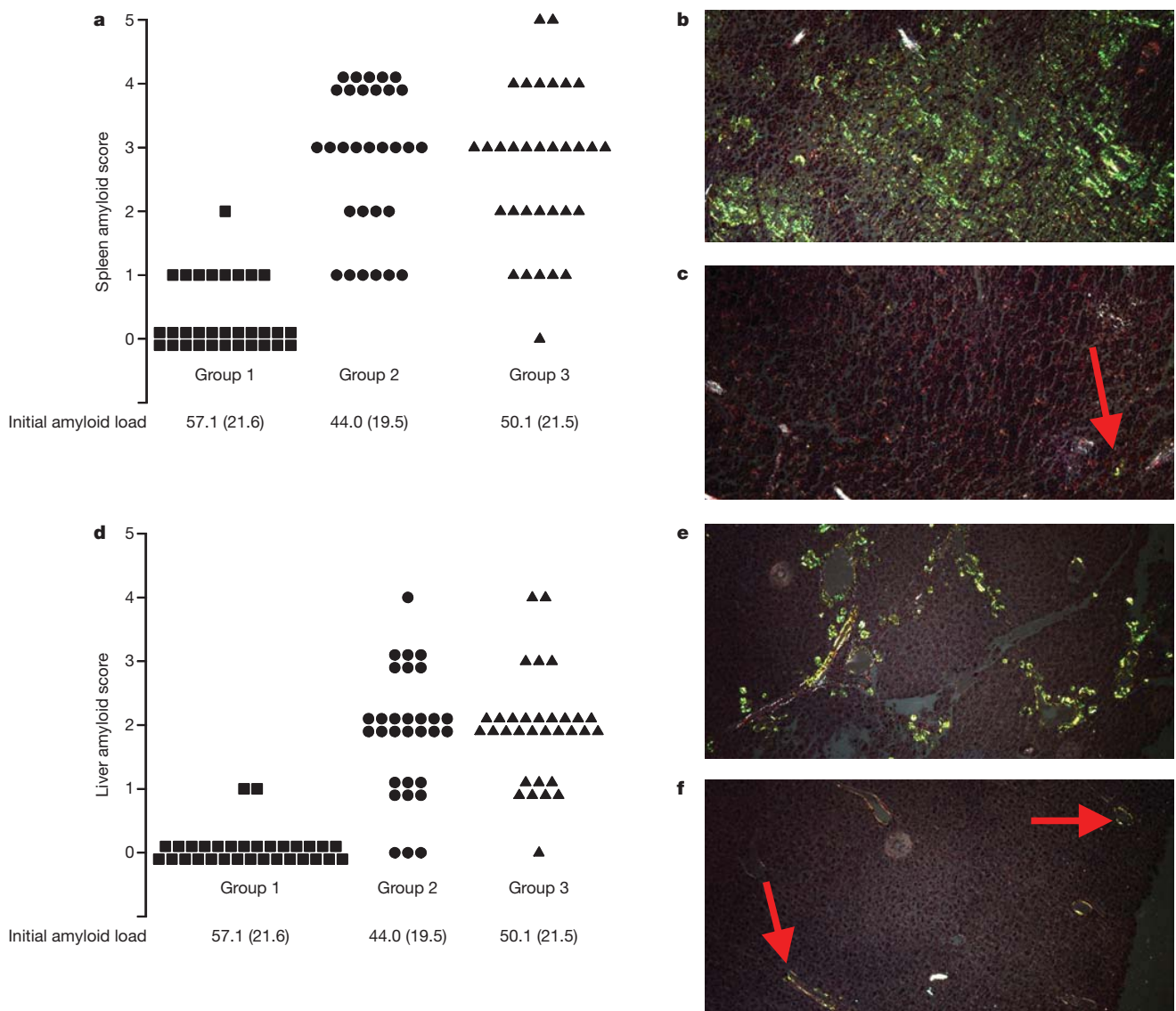


Figure 1 | Elimination of visceral amyloid in AA amyloidotic mice after treatment with anti-SAP antibody. **a–f**, Systemic AA amyloidosis was established in C57BL/6 SAP-deficient mice transgenically expressing human SAP, which were allocated to three groups closely matched for age, sex and whole-body amyloid load, the latter shown as mean (standard deviation) per cent whole-body retention of ^{125}I -human SAP tracer (one-way ANOVA comparing the three group means, $P = 0.054$). The mice were treated with CPHPC and a single dose of sheep anti-human-SAP antibody (group 1, $n = 31$), with CPHPC and the same dose of an irrelevant sheep anti-human antibody (group 2, $n = 30$), or left untreated (group 3, $n = 32$). Amyloid load was determined histologically 28 days later. **a**, Each point is an individual spleen amyloid score. 0, no amyloid detected; 1, one or more trace specks; 2, marginal zone traces; 3, general marginal zone deposits; 4, heavy marginal zone deposits; 5, heavy marginal zone and extensive interfollicular deposits. Kruskal–Wallis

residual amyloid was mostly being degraded within the cytoplasm of decreasing numbers of giant cells. Amyloid clearance was largely complete by about day 16 with remarkable restoration of normal tissue architecture and absence of any residual cellular infiltrate (Fig. 2o, s).

Human SAP binds avidly to mouse AA deposits *in vivo* and persists there with a half-life of 3–4 days, whereas circulating human SAP is cleared in mice with a half-life of 3–4 h and is undetectable in the plasma after 3 days^{4,10}. Amyloid deposits in non-transgenic AA amyloidotic C57BL/6 mice were thus loaded with human SAP by a single intraperitoneal injection of 10 mg of the isolated pure protein and anti-human-SAP antibody was injected 3 days later without the need for CPHPC. The same highly reproducible amyloid elimination occurred

test comparing the 3 groups, $P < 0.0001$. Dunn's multiple comparison test: 1 versus 2, $P < 0.001$; 1 versus 3, $P < 0.001$; 2 versus 3, not significant, $P > 0.05$. **b**, Congo-red-stained spleen section showing the pathognomonic amyloid green dichroism, score = 5. **c**, As in **b** but with amyloid score = 1; single amyloid speck is arrowed. **d**, Individual liver amyloid scores. 0, no amyloid detected; 1, trace specks; 2, traces in/around most portal tracts; 3, significant deposits in/around all portal tracts; 4, extensive portal and parenchymal deposits. Kruskal–Wallis test, $P < 0.0001$. Dunn's multiple comparison test: 1 versus 2, $P < 0.001$; 1 versus 3, $P < 0.001$; 2 versus 3, not significant, $P > 0.05$. There were no significant differences in the average spleen or liver amyloid scores between males and females within any of the groups (not shown). **e**, Congo-red stain of liver amyloid, score = 4. **f**, Liver amyloid score = 1; arrows indicate amyloid specks. All tissue micrographs are at $\times 100$ magnification.

as in the human SAP transgenic mice and this approach facilitated analysis of the mechanisms responsible.

In contrast to the clearance of amyloid deposits in wild-type mice, significantly more amyloid remained after anti-SAP treatment of complement-deficient animals lacking either C1q (ref. 12) or C3 (ref. 13) (Supplementary Information, section 7), demonstrating that the antibody effect is largely complement dependent. IgG antibody alone could potentially engage phagocytic cells via their Fc γ receptors and, although amyloid clearance was much reduced in the absence of complement, the persistent deposits in complement-deficient mice were more fragmented than in untreated controls, indicating some direct antibody effect. There was more complete amyloid elimination

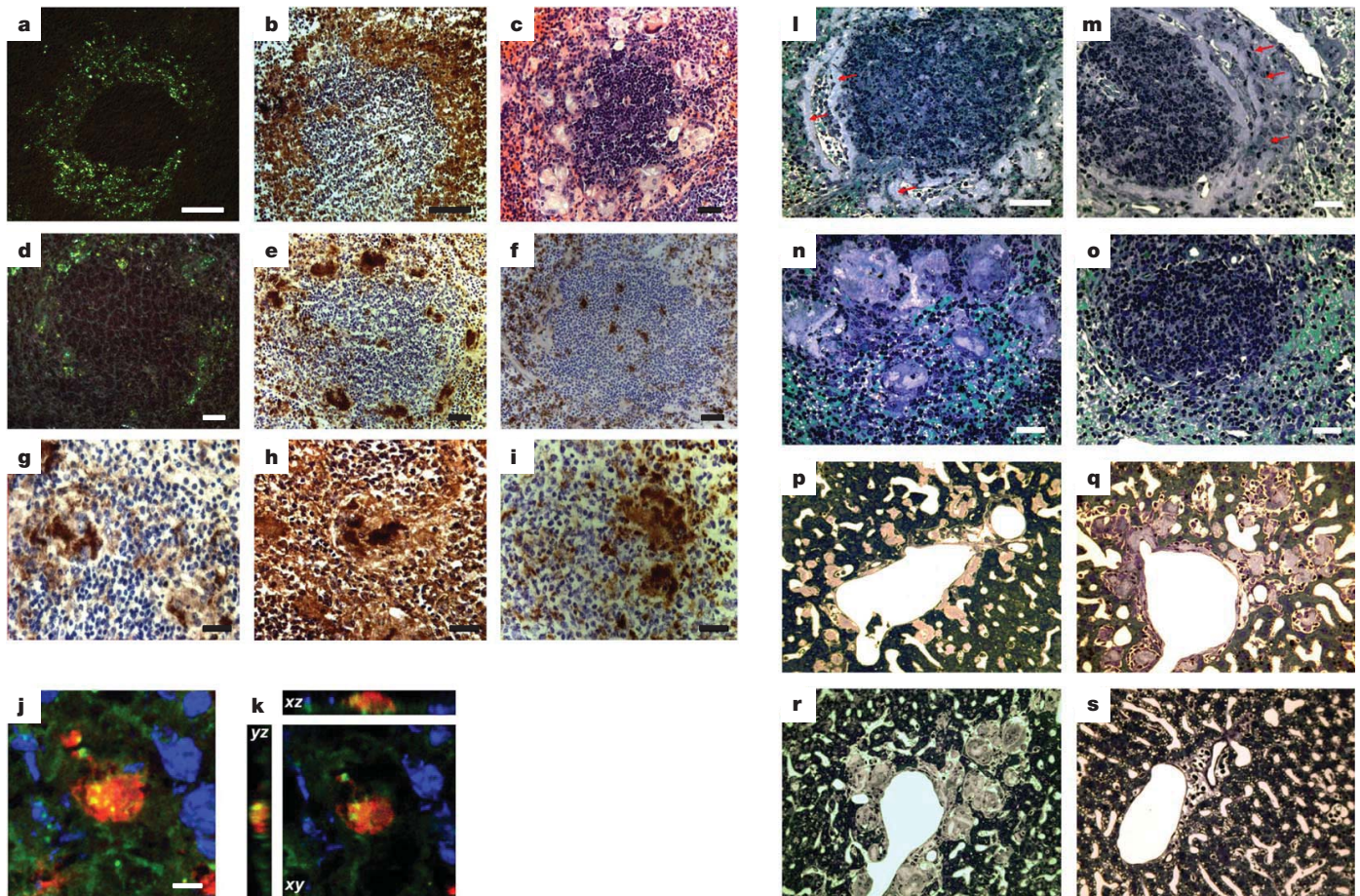


Figure 2 | Cellular infiltration and amyloid destruction after administration of anti-SAP antibody. **a**, Congo-red-stained spleen section one day after anti-SAP-antibody treatment showing typical heavy marginal zone amyloid. **b**, Anti-F4/80-stained adjacent section showing intense infiltration of the amyloid deposits with phagocytic cells. **c**, Haematoxylin-and-eosin-stained spleen section 4 days after anti-SAP-antibody treatment, showing multiple multinucleate giant cells surrounding and engulfing amyloid in the marginal zone. **d**, Congo-red-stained adjacent section showing marked fragmentation and reduction of marginal zone amyloid. **e**, Anti-CD68-stained adjacent section, showing massive infiltration of the marginal zone amyloid by phagocytically active macrophages and giant cells. **f**, Anti-CD68-stained spleen section from a control, untreated amyloidotic mouse, showing no positive cells in the pale amorphous marginal zone amyloid deposits. Scale bars in **a–f** are 100 μm . **g**, Splenic marginal zone amyloid deposit from a mouse 4 days after treatment with anti-SAP antibody, stained with antibodies to mouse AA, the amyloid fibril protein. **h**, Adjacent section stained with anti-mouse complement component C3. **i**, Adjacent section stained with anti-mouse CD68. Scale bars in **g–i** are 20 μm . **j**, Extended focus confocal view (Z projection) of the same spleen immunostained for CD68 (red), human serum

in some C1q-deficient mice than in C3-deficient mice (Supplementary Information, section 7), indicating that complement activation may occur in the absence of C1q but that C3 is critical. Consistent with this observation, F(ab)₂ anti-SAP-antibody treatment reduced amyloid load but was significantly less effective than intact IgG antibody (Supplementary Information, section 8). F(ab)₂ antibodies activate the alternative pathway, independently of C1q, and it is likely that the high dose of F(ab)₂ that was used (Supplementary Information, section 8) triggered some complement activation. Full efficacy of the anti-SAP antibody thus requires the Fc region but cellular recognition by Fc γ receptors is not a major factor as F(ab)₂ was more effective in complement-sufficient mice than IgG antibody was in complement-deficient mice.

When macrophage activity was ablated using liposomal clodronate¹⁴, anti-SAP antibody produced no reduction of amyloid load

amyloid A protein (SAA) (green) and counterstained with Hoechst 33342 (blue), confirming the close co-localization of amyloid and active phagolysosome fusion in macrophages and giant cells as they ingest and destroy the amyloid deposits that have been opsonised by anti-SAP antibody and complement. Scale bar, 5 μm . **k**, Orthogonal views of the same stain as **j**, showing ingested amyloid within a macrophage. **l–o**, **p–s**, Spleen (**l–o**; scale bars, 20 μm) and liver (**p–s**; magnification, $\times 250$), thin sections from mice in this experiment stained with toluidine blue. Control mice, not treated with anti-SAP antibody, show abundant amorphous amyloid deposits, pale blue in spleen (red arrows, **l**) and pink in liver (**p**) with the characteristic absence of any surrounding inflammatory reaction or cellular infiltrate. **m**, **q**, One day after anti-SAP antibody treatment showing intense, predominantly mononuclear cell infiltration (**m**, red arrows) in and around the amyloid. **n**, **r**, Five days after anti-SAP-antibody treatment showing fusion of macrophages to form multinucleate giant cells surrounding and infiltrating the deposits and containing large masses of ingested amyloid undergoing degradation. **o**, **s**, Sixteen days after anti-SAP-antibody administration showing complete elimination of amyloid deposits with no residual cellular infiltrate and restoration of normal tissue architecture.

(Supplementary Information, section 9), demonstrating that macrophages were the essential final effectors of amyloid clearance. Macrophages are largely responsible for the normal, clinically silent resolution of non-infective tissue injury and for remodelling of non-cellular matrix. The failure to spontaneously clear amyloid deposits—which are composed only of autologous constituents—is therefore remarkable, especially as, despite their inherent stability, amyloid fibrils can be digested by proteinases and phagocytic cells *in vitro*⁵, particularly when opsonised by antibody¹⁵. *In vivo* macrophage responses to different types of amyloid have been reported occasionally^{16–19}, and amyloid deposits sometimes regress when fibril-precursor-protein abundance is sufficiently reduced^{20,21}. However, amyloid usually accumulates with little or no local cellular or systemic inflammatory response. The serendipitous effect of CPHPC in depleting circulating SAP but leaving some SAP in amyloid deposits enabled the present use of

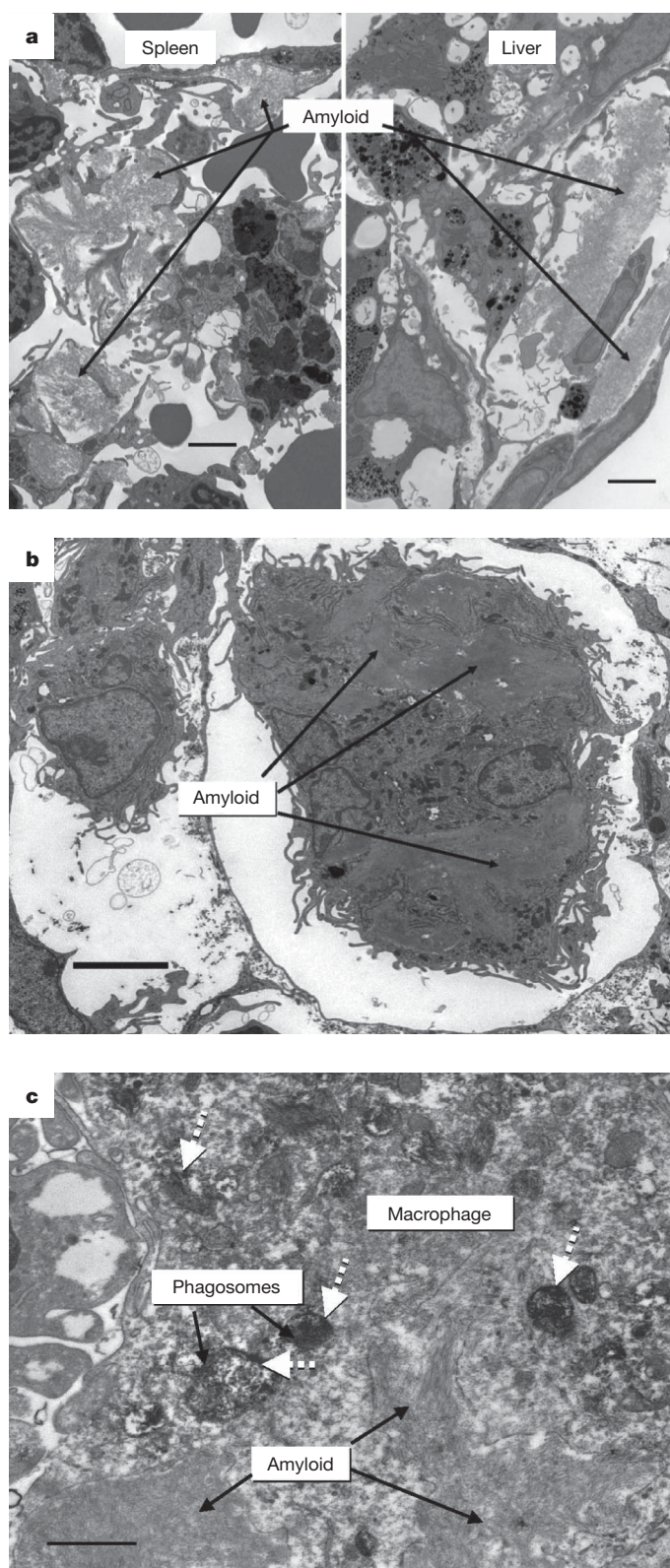


Figure 3 | Electron micrographs of amyloid destruction after anti-SAP-antibody treatment. **a**, Spleen and liver from a control AA amyloidotic mouse loaded with human SAP, which did not receive anti-SAP antibody, showing extracellular masses of fibrillar amyloid with the characteristic absence of any inflammatory cells or cellular reaction. Scale bars, 3 μ m. **b**, Liver from an AA amyloidotic mouse loaded with human SAP, examined 5 days after administration of anti-SAP antibody, showing a multinucleate giant cell surrounding, internalizing and digesting large masses of amyloid. Scale bar, 5 μ m. **c**, Anti-CD68 immunostain of spleen removed one day after administration of anti-SAP antibody to an AA amyloidotic mouse loaded with human SAP, showing a macrophage surrounding, infiltrating and ingesting amyloid. The dark peroxidase reaction product (white arrows) identifies phagolysosome fusion around internalized fragments of amyloid. Scale bar, 1 μ m.

produced by the sheep polyclonal anti-human SAP, and which had similar plasma half lives of ~ 4 days in wild-type C57BL/6 mice. IgG2a antibodies were selected because mouse IgG1 activates mouse complement poorly if at all²². SAP-5 and Abp1 recognized different epitopes on human SAP (Supplementary Information, section 10) but were each as potent as the polyclonal sheep anti-SAP in eliminating amyloid *in vivo* (Supplementary Information, sections 11 and 12).

Anti-SAP antibody could potentially elicit tissue-damaging inflammation in amyloidotic tissues. However, the present notable absence of any adverse effects presumably reflects the physiological nature of the macrophage reaction and is encouraging for the clinical use of CPHPC and anti-SAP. Nevertheless, appropriate caution will be essential because systemic amyloidosis patients have widespread amyloid deposits in sensitive tissues, including the heart, blood-vessel walls and nerves, which are not involved in the mouse AA model. Also, the trace amount of human SAP in normal glomerular basement membrane⁸ and elastic fibre microfibrils⁹ is a potential undesirable target for anti-SAP antibodies. It is therefore reassuring that there was no change in plasma biochemistry or any histological abnormality in human SAP transgenic mice treated with CPHPC followed by anti-human SAP antibodies (Supplementary Information, section 13).

Antibodies to amyloid β -protein (A β) are under intense investigation for the treatment of Alzheimer's disease and an *in vivo* imaging study²³ has shown binding to some human systemic immunoglobulin-light-chain amyloid (AL) deposits by a monoclonal anti-light-chain antibody that produces clearance of artefactual local human AL amyloidomas in mice^{24,25}. However, therapeutic anti-fibril antibodies will have to be reactive with each different type of amyloid whereas anti-SAP-antibody treatment is applicable to all forms of amyloidosis and all human amyloid deposits. Because the SAP that is universal in amyloid is derived from the circulation, anti-SAP antibodies and complement proteins will also be able to reach the deposits, and macrophages are present in, or can access, all tissues. Management of systemic amyloidosis will always require maximum efforts to reduce amyloid-fibril-precursor-protein production, if that is feasible, but the capacity to eliminate existing amyloid deposits would be a major therapeutic advance. We are now working towards clinical evaluation of this approach, and a candidate monoclonal anti-SAP antibody has been fully humanized for the exploration of safety, efficacy and optimal clinical dosing.

METHODS SUMMARY

Induction of murine AA amyloidosis using amyloid enhancing factor and repeated casein injections, estimation of amyloid load *in vivo* and *in vitro*, and quantification of human SAP in serum and tissue extracts, were conducted as previously reported^{4,6,10}. Sheep and mouse anti-human-SAP antibodies were raised by immunization with isolated pure human SAP²⁶ and mouse anti-human-SAP hybridomas were cloned by standard methods; Abp1 was produced by AbPro.

Received 29 June; accepted 13 September 2010.

Published online 20 October 2010.

anti-SAP antibodies to trigger unprecedented, clinically silent elimination of visceral amyloid deposits by macrophages.

The same therapeutic approach should be effective in human amyloidosis, using human or humanized monoclonal antibodies or other antibody constructs. Therefore we investigated two of our mouse monoclonal IgG2a anti-SAP antibodies, designated SAP-5 and Abp1, which bound to human SAP with similar affinities, on rates and off rates (Supplementary Information, section 10), which activated mouse complement *in vitro* producing C3 cleavage comparable to that

1. Pepys, M. B. Amyloidosis. *Annu. Rev. Med.* **57**, 223–241 (2006).
2. Pepys, M. B., Dyck, R. F., de Beer, F. C., Skinner, M. & Cohen, A. S. Binding of serum amyloid P component (SAP) by amyloid fibrils. *Clin. Exp. Immunol.* **38**, 284–293 (1979).
3. Pepys, M. B. *et al.* Amyloid P component. A critical review. *Amyloid* **4**, 274–295 (1997).

4. Pepys, M. B. *et al.* Targeted pharmacological depletion of serum amyloid P component for treatment of human amyloidosis. *Nature* **417**, 254–259 (2002).
5. Tennent, G. A., Lovat, L. B. & Pepys, M. B. Serum amyloid P component prevents proteolysis of the amyloid fibrils of Alzheimer's disease and systemic amyloidosis. *Proc. Natl Acad. Sci. USA* **92**, 4299–4303 (1995).
6. Botto, M. *et al.* Amyloid deposition is delayed in mice with targeted deletion of the serum amyloid P component gene. *Nature Med.* **3**, 855–859 (1997).
7. Gillmore, J. D. *et al.* Sustained pharmacological depletion of serum amyloid P component in patients with systemic amyloidosis. *Br. J. Haematol.* **148**, 760–767 (2010).
8. Dyck, R. F. *et al.* Amyloid P-component is a constituent of normal human glomerular basement membrane. *J. Exp. Med.* **152**, 1162–1174 (1980).
9. Breathnach, S. M. *et al.* Amyloid P component is located on elastic fibre microfibrils of normal human tissues. *Nature* **293**, 652–654 (1981).
10. Hawkins, P. N., Myers, M. J., Epenetos, A. A., Caspi, D. & Pepys, M. B. Specific localization and imaging of amyloid deposits *in vivo* using ¹²⁵I-labeled serum amyloid P component. *J. Exp. Med.* **167**, 903–913 (1988).
11. Hawkins, P. N. & Pepys, M. B. A primed state exists *in vivo* following histological regression of amyloidosis. *Clin. Exp. Immunol.* **81**, 325–328 (1990).
12. Botto, M. *et al.* Homozygous C1q deficiency causes glomerulonephritis associated with multiple apoptotic bodies. *Nature Genet.* **19**, 56–59 (1998).
13. Wessels, M. R. *et al.* Studies of group B streptococcal infection in mice deficient in complement component C3 or C4 demonstrate an essential role for complement in both innate and acquired immunity. *Proc. Natl Acad. Sci. USA* **92**, 11490–11494 (1995).
14. Van Rooijen, N. & Hendriks, E. Liposomes for specific depletion of macrophages from organs and tissues. *Methods Molec. Biol.* **605**, 189–203 (2009).
15. Zucker-Franklin, D. Immunophagocytosis of human amyloid fibrils by leukocytes. *J. Ultrastruct. Res.* **32**, 247–257 (1970).
16. Argilés, À., García García, M. & Mourad, G. Phagocytosis of dialysis-related amyloid deposits by macrophages. *Nephrol. Dial. Transplant.* **17**, 1136–1138 (2002).
17. Nakamura, Y. *et al.* A case of hereditary amyloidosis transthyretin variant Met 30 with amyloid cardiomyopathy, less polyneuropathy, and the presence of giant cells. *Pathol. Int.* **49**, 898–902 (1999).
18. Verine, J. *et al.* Clinical and histological characteristics of renal AA amyloidosis: a retrospective study of 68 cases with a special interest to amyloid-associated inflammatory response. *Hum. Pathol.* **38**, 1798–1809 (2007).
19. Weiss, S. W. & Page, D. L. Amyloid nephropathy of Osterlag with special reference to renal glomerular giant cells. *Am. J. Pathol.* **72**, 447–455 (1973).
20. Lachmann, H. J. *et al.* Outcome in systemic AL amyloidosis in relation to changes in concentration of circulating free immunoglobulin light chains following chemotherapy. *Br. J. Haematol.* **122**, 78–84 (2003).
21. Lachmann, H. J. *et al.* Natural history and outcome in systemic AA amyloidosis. *N. Engl. J. Med.* **356**, 2361–2371 (2007).
22. Klaus, G. G. B., Pepys, M. B., Kitajima, K. & Askonas, B. A. Activation of mouse complement by different classes of mouse antibody. *Immunology* **38**, 687–695 (1979).
23. Wall, J. S. *et al.* Radioimmunodetection of amyloid deposits in patients with AL amyloidosis. *Blood*. doi:10.1182/blood-2010-03-273797 (3 June 2010).
24. Solomon, A., Weiss, D. T. & Wall, J. S. Immunotherapy in systemic primary (AL) amyloidosis using amyloid-reactive monoclonal antibodies. *Cancer Biother. Radiopharm.* **18**, 853–860 (2003).
25. Solomon, A., Weiss, D. T. & Wall, J. S. Therapeutic potential of chimeric amyloid-reactive monoclonal antibody 11–1F4. *Clin. Cancer Res.* **9**, 3831S–3838S (2003).
26. de Beer, F. C. & Pepys, M. B. Isolation of human C-reactive protein and serum amyloid P component. *J. Immunol. Methods* **50**, 17–31 (1982).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The study was supported by Medical Research Council Programme Grant G97900510 to M.B.P. and P.N.H. and by the University College London Amyloidosis Research Fund. K.B. was supported by the Erik and Edith Fernströms Foundation for Medical Research and a Postdoctoral Fellowship from the Swedish Research Council. K.B. dedicates his work to his late father, Gunnar Bodin. We thank S. Gordon, S. Wood, S. Kolstoe, J. Raynes, P. Simons and R. Al-Shawi for information, advice, reagents and support, and B. Jones for processing the manuscript.

Author Contributions The study was conceived, designed and supervised by M.B.P. K.B., M.C.K. and S.E. performed all the experimental animal work. G.A.T., A.L., J.A.G., S.M. and A.P.D. performed or contributed to the histological studies. Amyloid scoring was performed by K.B., M.C.K., S.E., J.D.G. and M.B.P. W.L.H., P.P.M., J.R.G., D.J.M., G.W.T. and V.B. conducted the immunochemical, radiochemical and immunoassay studies. A.P. undertook the statistical analyses. A.R.B. produced the sheep anti-human SAP and control antisera. M.B. supplied the complement knockout mice. J.D.G. and P.N.H. contributed to experimental design. The paper was written by M.B.P. and reviewed and approved by all co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.B.P. (m.pepys@ucl.ac.uk).

RANK ligand mediates progesterin-induced mammary epithelial proliferation and carcinogenesis

Eva Gonzalez-Suarez^{1*†}, Allison P. Jacob^{1*}, Jon Jones¹, Robert Miller¹, Martine P. Roudier-Meyer², Ryan Erwert¹, Jan Pinkas^{3†}, Dan Branstetter² & William C. Dougall¹

RANK ligand (RANKL), a TNF-related molecule, is essential for osteoclast formation, function and survival through interaction with its receptor RANK^{1,2}. Mammary glands of RANK- and RANKL-deficient mice develop normally during sexual maturation, but fail to form lobuloalveolar structures during pregnancy because of defective proliferation and increased apoptosis of mammary epithelium³. It has been shown that RANKL is responsible for the major proliferative response of mouse mammary epithelium to progesterone during mammary lactational morphogenesis⁴, and in mouse models, manipulated to induce activation of the RANK/RANKL pathway in the absence of strict hormonal control, inappropriate mammary proliferation is observed^{5,6}. However, there is no evidence so far of a functional contribution of RANKL to tumorigenesis. Here we show that RANK and RANKL are expressed within normal, pre-malignant and neoplastic mammary epithelium, and using complementary gain-of-function (mouse mammary tumour virus (MMTV)-RANK transgenic mice) and loss-of function (pharmacological inhibition of RANKL) approaches, define a direct contribution of this pathway in mammary tumorigenesis. Accelerated pre-neoplasias and increased mammary tumour formation were observed in MMTV-RANK transgenic mice after multiparity or treatment with carcinogen and hormone (progesterone). Reciprocally, selective pharmacological inhibition of RANKL attenuated mammary tumour development not only in hormone- and carcinogen-treated MMTV-RANK and wild-type mice, but also in the MMTV-*neu* transgenic spontaneous tumour model. The reduction in tumorigenesis upon RANKL inhibition was preceded by a reduction in pre-neoplasias as well as rapid and sustained reductions in hormone- and carcinogen-induced mammary epithelial proliferation and cyclin D1 levels. Collectively, our results indicate that RANKL inhibition is acting directly on hormone-induced mammary epithelium at early stages in tumorigenesis, and the permissive contribution of progesterone to increased mammary cancer incidence is due to RANKL-dependent proliferative changes in the mammary epithelium. The current study highlights a potential role for RANKL inhibition in the management of proliferative breast disease.

Previously we reported that MMTV-RANK female transgenic mice show sustained proliferation and impaired alveolar secretory differentiation of the mammary epithelium upon pregnancy; however, spontaneous mammary tumours were not observed in aged virgin mice⁶. In contrast, mammary glands from multiparous MMTV-RANK mice showed spontaneous tumour development (median onset 26.5 months) and exhibited a higher incidence of pre-neoplasias compared with wild-type mice; adenocarcinomas were only observed in MMTV-RANK mice (Supplementary Table 1).

To characterize more fully the role of the RANKL/RANK pathway in mammary tumorigenesis, we induced mammary carcinogenesis using

combined treatment with the hormone medroxyprogesterone acetate (MPA) and a carcinogen (7,12-dimethylbenzanthracene (DMBA)) (Supplementary Fig. 1)⁷. After MPA/DMBA treatment, MMTV-RANK transgenic mice showed a markedly enhanced susceptibility to mammary tumours compared with wild-type mice (Fig. 1a). Adenocarcinoma, adenosquamous carcinoma and adenomyoepithelioma carcinoma histotypes were observed in both strains (Supplementary Fig. 2). Pre-neoplastic mammary lesions were clearly more widely distributed in mammary tissues in MMTV-RANK mice than in wild-type mice (Fig. 1b). Multifocal ductal hyperplasias, multifocal and focally extensive mammary intraepithelial neoplasias (MIN), and multiple carcinomas were frequently present in a single involved MMTV-RANK mammary gland in contrast with focal lesions in the wild-type glands. Whole-mount analysis of mammary glands at a stage before detection of palpable tumours revealed a 100% incidence and high multiplicity of dense pre-neoplastic epithelial foci, consisting of a mix of hyperplasias and MIN, in MMTV-RANK mice (Fig. 1c). In the induced model, increased mammary proliferation was also evident in the MMTV-RANK versus wild-type glands as early as 2 days after the first DMBA treatment, and a significant increase in cyclin D1 immunohistochemistry signal was evident at 4 and 7 weeks after the last DMBA treatment (Fig. 1d).

RANK protein was clearly evident in luminal and abluminal cells of normal mammary epithelium and in the epithelial component of pre-neoplasias and carcinomas in both MPA/DMBA-treated wild-type and MMTV-RANK mice, and as expected, RANK was observed at higher levels in mammary tissue of MMTV-RANK mice (Supplementary Figs 3 and 4). Compared with ductal hyperplasias and adenocarcinomas, RANK expression was relatively lower in MIN lesions from both wild-type and MMTV-RANK mice (Supplementary Figs 3 and 4). In mammary glands from tumour-bearing wild-type and MMTV-RANK mice, RANKL protein was detected within the luminal cells of histopathologically normal mammary epithelium and in the epithelial component of pre-malignant lesions, with consistently higher levels detected in MIN lesions as compared to other pre-malignant tissues (Supplementary Figs 3 and 4). RANKL was also detected within the carcinoma element in the majority of each histotype in wild-type and MMTV-RANK tumours. The lower level of RANK immunoreactivity within MIN may be explained by a ligand-dependent reduction of RANK protein or mRNA, as demonstrated in mammary tissues of MMTV-RANKL mice⁵ or after treatment of mammary epithelial cells (MECs) with RANKL *in vitro*⁶, respectively.

Given the potential importance of the RANK/RANKL pathway early in tumorigenesis, we confirmed that MPA pre-treatment resulted in an early and marked induction of RANKL (Supplementary Fig. 5). Expression of both RANK and RANKL proteins was also evident at the time of carcinogen administration (Supplementary Fig. 5). Dual immunostaining of samples from MPA/DMBA-treated mice indicated

¹Department of Hematology/Oncology Research, Amgen Inc, Seattle, Washington 98119, USA. ²Department of Pathology, Amgen Inc, Seattle, Washington 98119, USA. ³Department of Hematology/Oncology Research, Amgen Inc, Cambridge, Massachusetts 02139, USA. [†]Present addresses: Cancer Epigenetics and Biology Program, Bellvitge Institute for Biomedical Research, 08907-L'Hospitalet de Llobregat-Barcelona, Spain (E.G.-S.); ImmunoGen Inc, Waltham, Massachusetts 02451, USA (J.P.).

*These authors contributed equally to this work.

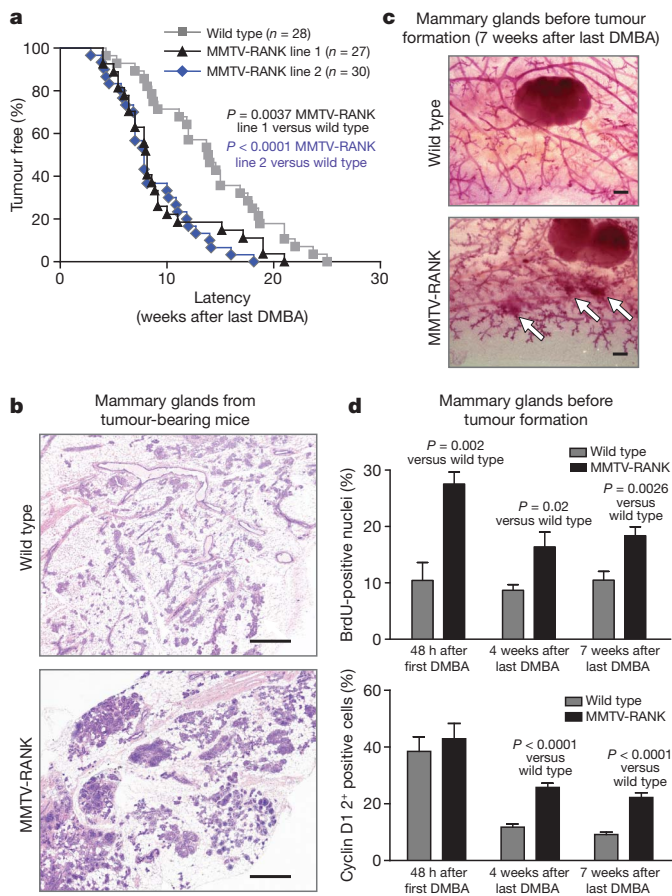


Figure 1 | Decreased latency of mammary tumour formation and accelerated pre-neoplastic changes in mammary glands of MMTV-RANK mice treated with hormone and carcinogen. **a**, Kinetics of palpable mammary tumour onset in wild-type and MMTV-RANK lines 1 and 2 mice treated with MPA plus DMBA; P -values based on log-rank test. Latency was expressed as weeks after last DMBA treatment. **b**, Mammary glands from tumour-bearing wild-type and MMTV-RANK line 1 mice after MPA plus DMBA treatment (haematoxylin and eosin staining; original magnification, $\times 4$). Scale bar, 500 μm . **c**, **d**, Analysis of mammary glands before tumour formation. **c**, Whole mounts of inguinal mammary glands from wild-type and MMTV-RANK line 1 mice prepared 7 weeks after last DMBA treatment; arrows indicate dense epithelial foci. Scale bar, 500 μm . **d**, Proliferation in mammary epithelial cells (BrdU immunohistochemistry, mean \pm s.e.m.) and quantification of high cyclin D1 expressing cells (2+ score by immunohistochemistry) in glands from MPA/DMBA-treated wild-type and MMTV-RANK line 1 mice 48 h after the first DMBA treatment and 4 and 7 weeks after the last DMBA treatment. $n = 5$ mice per group. Independent t -tests were used for statistical comparisons.

that RANKL protein was predominately localized in progesterone receptor (PR)-expressing cells, not only in normal luminal epithelial cells (as has been described)⁵ but also in pre-neoplastic and neoplastic epithelium (Fig. 2a–d). Additionally, some RANKL immunoreactivity was also detected in luminal spaces, perhaps reflecting secreted material or induction via other factors such as prolactin or parathyroid hormone-related peptide³. In human invasive breast carcinoma, RANKL was detected heterogeneously in 11% (6 of 57) of tumours (Fig. 2e) whereas RANK was detected in 6% (3 of 50) of tumours (Fig. 2h). RANK and RANKL were not co-expressed within the carcinoma epithelium of any sample. Unlike the mouse model, RANKL was additionally detected in sporadic infiltrating mononuclear cells present in most tumour stroma (67%; 38 of 57) and in fibroblast-like cells in the stroma of rare tumours (5%; 3 of 57) (Fig. 2f and g, respectively).

To elucidate the operative role of RANKL in mammary tumorigenesis, MPA/DMBA-induced MMTV-RANK mice were treated with a selective pharmacological RANKL inhibitor, RANK-Fc, concurrently with the

first DMBA treatment, but 3 weeks after MPA treatment. RANK-Fc significantly delayed formation of palpable mammary tumours (Fig. 3a). Whereas adenocarcinomas were the predominant carcinoma subtype observed in MPA/DMBA-induced tumours in MMTV-RANK mice, RANK-Fc significantly suppressed adenocarcinoma formation such that the spectrum of carcinoma histotypes in mice that eventually developed tumours was evenly distributed (Supplementary Fig. 6). Because we had also observed RANKL and RANK expression in malignant and pre-malignant lesions in wild-type mice, we also tested whether RANKL affected tumour formation in MPA/DMBA-treated wild-type mice. RANK-Fc, initiated concurrently with carcinogen administration, almost completely blocked the occurrence of palpable mammary tumours in wild-type mice (Fig. 3b). The reduced incidence and delay in mammary tumour formation in MPA/DMBA-induced wild-type mice was observed selectively with RANK-Fc treatment, but not after treatment with the bisphosphonate zoledronic acid (Supplementary Fig. 7). The number of F4/80⁺ myeloid cells in the tumour micro-environment was not altered after RANK-Fc treatment (data not shown). Altogether these data indicate that RANKL inhibition blocks mammary tumour formation via a direct effect on mammary epithelia, distinct from its established ability to inhibit osteoclastogenesis.

Analysis of mammary glands from tumour-bearing MMTV-RANK and wild-type mice showed that RANKL inhibition led to a selective reduction in the proliferative index of resident normal mammary epithelium and pre-neoplastic hyperplasias, but not in adenocarcinomas (Fig. 3c), and increased apoptosis was observed in MIN (Fig. 3d). These alterations manifested into a marked reduction in overall mammary epithelial density, as evident in either MMTV-RANK or wild-type mice treated with RANK-Fc (Supplementary Fig. 8). Because our data suggested direct effects on mammary epithelium and protection against early events in tumorigenesis, we analysed pre-neoplasias, cyclin D1 levels^{8,9} and proliferation in mice at earlier stages, before the detection of palpable mammary tumours. Treatment of MPA/DMBA-induced MMTV-RANK mice with RANK-Fc resulted in a clear suppression of pre-neoplastic lesions and a corresponding reduction in mammary epithelial proliferation at early stages (Supplementary Fig. 8 and Supplementary Table 2). Elevated cyclin D1 has been reported during RANKL-dependent hyperplasia⁵ and RANKL-dependent normal morphological changes during mammary gland development¹⁰; we also demonstrated that RANKL treatment causes increased proliferation⁶ and rapidly increases cyclin D1 mRNA and protein levels in MMTV-RANK MECs *in vitro* (Supplementary Fig. 9), indicating that proliferation and cyclin D1 were potential pharmacodynamic markers of RANK-Fc activity in this tissue. RANK-Fc treatment of either MMTV-RANK or wild-type mice caused a very rapid reduction (within 48 h of treatment) in both hormone-induced proliferation and mammary epithelial cyclin D1 levels (Supplementary Fig. 8).

Progesterone triggers mammary proliferation via two distinct mechanisms: an early, RANKL-independent, direct mitogenic effect on PR⁺ cells followed by a wave of greater proliferation mediated by the paracrine effect of RANKL on PR[−] cells⁴. We observed that pre-treatment with MPA was sufficient to increase both RANKL and proliferation at time of subsequent DMBA exposure. These extremely early proliferative changes, due to either MPA pre-treatment alone or as a result of combined MPA and DMBA treatment, were dependent on RANKL as shown by the significant and rapid reductions upon RANK-Fc treatment (Supplementary Fig. 5). To address the relative contribution of RANKL to the amplification of early progesterone-induced proliferation, we compared the effect of a RANK-Fc dose which maximally inhibits proliferation (10 mg kg^{−1}; Supplementary Fig. 10) to the PR antagonist mifepristone. Injection of unsynchronized cycling wild-type mice with mifepristone resulted in a 98% reduction in progesterone-induced proliferation ($P < 0.001$ versus control) whereas RANK-Fc reduced proliferation by 82% ($P < 0.001$ versus control), indicating that RANKL was responsible for the majority of proliferation induced by progesterone at 72 h (Supplementary Fig. 10). Notably, we observed that

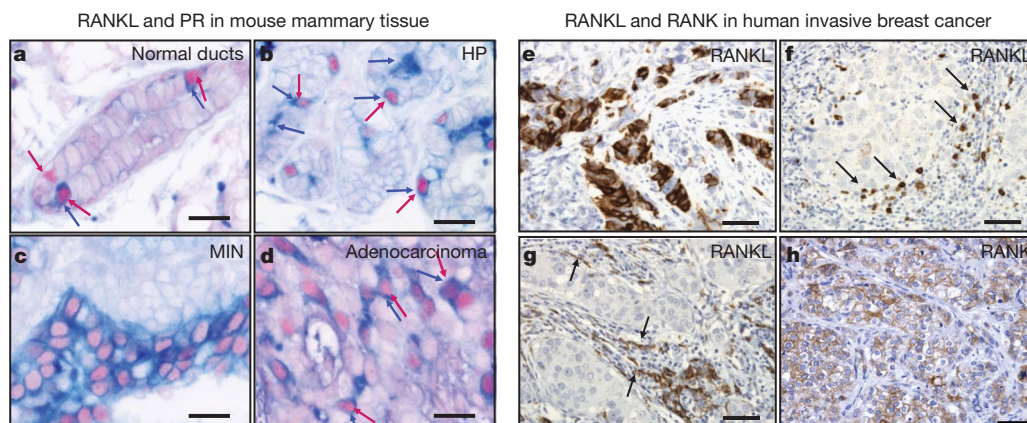


Figure 2 | RANKL is localized in PR⁺ cells at multiple stages during mouse tumorigenesis, and RANKL and RANK are expressed in invasive carcinoma of the human breast. **a–d**, Representative examples of immunohistochemistry analysis of RANKL and PR in mammary glands from tumour-bearing wild-type mice after MPA plus DMBA treatment showing normal mammary epithelium (**a**), hyperplasia (**b**), MIN (**c**) and adenocarcinoma (**d**); RANKL is blue and PR is red and indicated by arrows. RANKL protein was rarely observed within mononuclear cells in the stroma but was detected within the epithelial component of normal mammary epithelium, hyperplasia, MIN and

adenocarcinoma. RANKL expression is most frequently localized in PR-expressing cells although there is some RANKL immunoreactivity also detected in luminal spaces distinct from PR-expressing cells. Scale bar, 40 µm.

e–h, Representative examples of RANKL or RANK immunohistochemistry (brown) of human invasive breast carcinoma ($\times 20$ magnification). **e**, Intense RANKL staining within the tumour epithelium. **f**, Scattered mononuclear cells with intense RANKL reactivity. **g**, Intense RANKL reactivity was detected in spindle-shaped fibroblast-like stromal cells of some tumours. **h**, RANK was detected in epithelial carcinoma cells in some tumours. Scale bar, 50 µm.

RANK-Fc had no effect on RANKL levels whereas mifepristone completely blocked progesterone-induced RANKL production (Supplementary Fig. 10), indicating that RANKL inhibition was not directly interfering with progesterone-PR interactions. RANKL treatment of MECs from MMTV-RANK mice stimulated growth in semi-solid media,

caused an inappropriate filling of the luminal cavity⁶ and disrupted normal architecture and apicobasal polarization of three-dimensional acinar cultures (Supplementary Fig. 9), indicating that RANKL can also evoke phenotypic changes consistent with transformation.

We next examined the effect of RANKL inhibition in MMTV-*neu* transgenic mice, a spontaneous mammary tumour model without an exogenous hormone requirement¹¹. Analysis of virgin cycling MMTV-*neu* mice at a relatively late stage, but still before detection of palpable mammary tumours (4–5-month-old mice), revealed RANK and RANKL expression in normal mammary epithelium (Fig. 4a). RANK-Fc treatment, beginning at 5 months, had no significant impact on median time

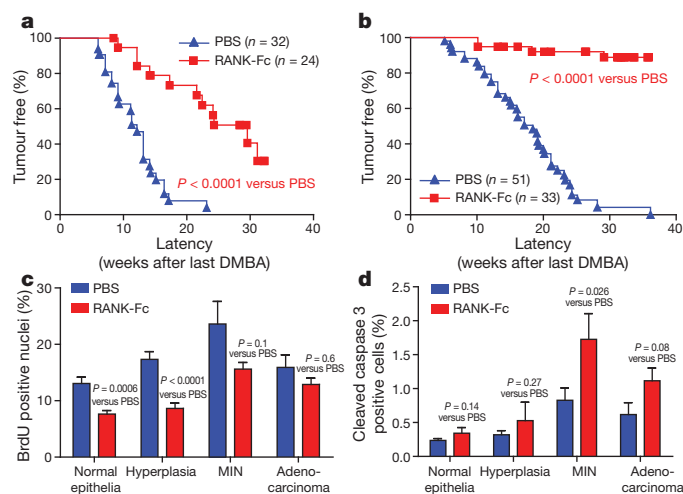


Figure 3 | RANK-Fc inhibits MPA/DMBA-induced mammary tumour formation which is preceded by a protective effect on early pre-neoplastic changes in the mammary gland. RANK-Fc treatment, at 10 mg kg⁻¹ subcutaneously three times per week, was initiated as indicated in Supplementary Fig. 1 and continued until mice were killed. **a**, Kinetics of palpable mammary tumour onset in MPA/DMBA-treated MMTV-RANK line 1 mice after treatment with PBS or RANK-Fc. **b**, Kinetics of palpable mammary tumour onset in MPA/DMBA-treated wild-type mice after treatment with PBS or RANK-Fc. Data for panels **a** and **b** are expressed as percentage of mice free of palpable tumours after the last DMBA treatment (*P*-values based on log-rank test). **c**, **d**, Distinct regions of interest (normal mammary epithelium, hyperplasia, MIN and adenocarcinoma) of mammary glands from tumour-bearing mice were identified histologically in ≥ 5 individual mice per group, circumscribed, and each distinct region of interest was separately assessed for proliferation (BrdU immunohistochemistry) or apoptosis (cleaved caspase 3 immunohistochemistry). Data for proliferation (**c**) and apoptosis (**d**) in mammary tissue from MMTV-RANK tumour-bearing mice with or without RANK-Fc treatment are expressed as mean per cent \pm s.e.m.; independent *t*-tests were used for statistical comparisons.

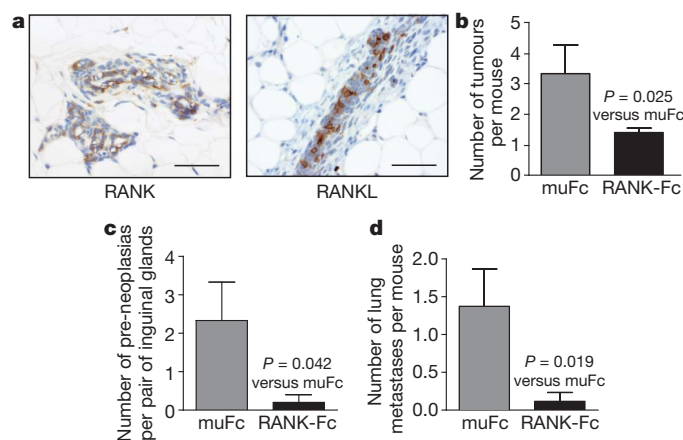


Figure 4 | RANK and RANKL are expressed in normal mammary epithelium of MMTV-*neu* mice, and RANK-Fc decreases spontaneous mammary tumorigenesis and lung metastasis in this model. **a**, RANK and RANKL expression (immunohistochemistry) in normal mammary epithelium is observed at a stage before palpable mammary tumour formation (4–5-month-old MMTV-*neu* mice). Scale bar, 100 µm. **b–d**, Effects of vehicle control (muFc) or RANK-Fc administered beginning in 5-month-old MMTV-*neu* mice (*n* = 10 per group). **b**, Mean number (\pm s.e.m.) of palpable tumours (all glands) plus tumours identified histologically (inguinal glands) in MMTV-*neu* mice with or without RANK-Fc. **c**, Mean number (\pm s.e.m.) of histologically defined pre-neoplasias in inguinal mammary glands of tumour-bearing MMTV-*neu* mice with or without RANK-Fc. **d**, Mean number of lung metastases (\pm s.e.m.) in tumour-bearing MMTV-*neu* mice with or without RANK-Fc. Entire lungs were step sectioned at 75 µm and individual metastases identified histologically. Independent *t*-tests were used for all statistical comparisons.

to spontaneous mammary tumour onset (Supplementary Fig. 11); however, a significant reduction in total number of mammary tumours quantified at necropsy was observed (Fig. 4b). RANKL inhibition again resulted in reduced pre-neoplasias (Fig. 4c), with correspondent reductions in proliferation and cyclin D1 selectively within normal mammary epithelia, but not in adenocarcinomas (Supplementary Fig. 11). Treatment of MMTV-*neu* mice with RANK-Fc also significantly reduced the incidence (Supplementary Fig. 11) and number of lung metastases per mouse (Fig. 4d). The reduced metastasis may reflect the reduced primary tumour burden after RANK-Fc treatment, but could also be due to inhibition of the metastasis suppressor Maspin by RANKL¹². Despite the lack of hormone receptors in MMTV-*neu* mammary tumours¹³ (Supplementary Fig. 12), ovarian hormones (including progesterone) can promote tumorigenesis in these mice selectively during a 'risk window', before 5 months of age¹⁴. Importantly, we observed that RANKL and PR are expressed in normal mammary epithelia but are not detected in any stage of atypia, including MIN and adenocarcinoma (Supplementary Fig. 12). Therefore, the demonstrated abilities of RANK-Fc to reduce mammary epithelial proliferation, protect against pre-neoplasias, and reduce tumour multiplicity in this spontaneous mammary tumour model are similar to the early stage effects observed in the induced MPA/DMBA model and could conceivably result from inhibition of repetitive cycles of RANKL expression in normal PR⁺ mammary glands of MMTV-*neu* mice, driven by progesterone surges at dioestrus¹⁵. The potential efficacy of RANKL inhibition at later stages or in hormone-receptor-negative tumours is currently under investigation.

Progesterone may accelerate mammary tumorigenesis in this and other rodent mammary tumour models^{7,16} via a proliferative expansion of the target cell pool available for carcinogenic damage¹⁷; however, defined molecular targets of progesterone function in tumorigenesis have not been identified. Recently, two studies^{15,18} revealed a critical involvement of RANKL in the expansion of mammary stem cells by progesterone, elevated either at pregnancy or cyclically, during the luteal dioestrus phase. It has been hypothesized that increased numbers and/or activity of mammary stem cells may predispose to mammary cancer by providing a greater target pool for transformation^{15,18}. This suggests that RANKL inhibition may also potentially mediate non-proliferative changes (for example, reductions in mammary stem cells) in addition to the reduced proliferation and delayed onset of pre-neoplasias observed in this study, which collectively reduces mammary tumorigenesis. Progesterone has been linked to the aetiology of human breast cancer due to its mitogenic effects¹⁹. The mammary epithelial proliferative index is highest during the luteal phase or in women receiving combined oestrogen plus progesterone hormone replacement therapy (HRT), as compared with women treated with oestrogen alone²⁰. Significant increases in breast density²¹, breast cancer risk²² and recurrence²³ have also been observed when progestins were included in HRT. Our data indicate that the pivotal role of progesterone in the promotion and growth of breast tumours may be mediated by RANKL-dependent effects. The observation of RANKL within human breast tumour stroma (that is, fibroblast-like cells and mononuclear cells), in addition to the tumour epithelium (this study and ref. 24), indicates that RANKL protein expression may not be modulated exclusively by hormone, and other mechanisms may exist to dysregulate and activate this pathway in breast cancer.

METHODS SUMMARY

Mouse experimental procedure. MMTV-RANK lines 1 and 2 were established (C57BL/6 background) and maintained as described⁶. Multiparous females were observed for the development of spontaneous tumours up to 29 months of age (>5 pregnancies for each animal). Induction of mammary tumours by MPA and DMBA were in accordance with the published protocol⁷. MMTV-RANK line 1 and wild-type C57BL/6 mice were treated with 10 mg kg⁻¹ murine RANK-Fc (Amgen Inc.) or PBS subcutaneously three times per week. Zoledronic acid (Zometa) was dosed at 0.5, 0.2, or 0.025 mg kg⁻¹ weekly. Treatment was initiated with the first DMBA treatment at 9 weeks of age and continued until mice were

killed. MMTV-*neu* (N202 Mul; FvB background) mice were obtained from Jackson Laboratories. 10 mg kg⁻¹ RANK-Fc or muFc (IgG1, both Amgen Inc.) were given three times per week beginning at 5 months of age and continued until mice were killed. Tumours were detected by manual palpation weekly.

Whole-mount analysis, histology and immunohistochemistry. For whole-mount analysis, both inguinal mammary glands were removed at the time specified (either 4 or 7 weeks after the last DMBA treatment), fixed and stained with carmine²⁵. Dense epithelial foci of greater than 300 µm were counted. For histological and immunohistochemistry analysis, tissues were fixed in formalin or Zinc-Tris solution and embedded in paraffin. Sections (4 µm) were stained with haematoxylin and eosin, and assessed according to the Annapolis guidelines²⁶. BrdU, cyclin D1, cleaved caspase 3, RANK and RANKL stained sections were visualized with diaminobenzidine and quantified using Aperio Spectrum software (Aperio Technologies). Dual RANKL- and PR-stained sections were visualized with alkaline phosphatase Ferangi blue (Biocare Medical) for RANKL and permanent red (Dako) for PR. Formalin-fixed, paraffin-embedded tumour specimens (57 samples of invasive carcinoma of human breast) were obtained (Asterand) and immunohistochemistry was performed with M366 (anti-huRANKL monoclonal antibody; Amgen Inc.) or N-1H8 and N-2B10 (anti-huRANK monoclonal antibodies; Amgen Inc.).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 November 2009; accepted 13 September 2010.

Published online 29 September 2010.

- Dougall, W. C. *et al.* RANK is essential for osteoclast and lymph node development. *Genes Dev.* **13**, 2412–2424 (1999).
- Kong, Y. Y. *et al.* OPG is a key regulator of osteoclastogenesis, lymphocyte development and lymph-node organogenesis. *Nature* **397**, 315–323 (1999).
- Fata, J. E. *et al.* The osteoclast differentiation factor osteoprotegerin-ligand is essential for mammary gland development. *Cell* **103**, 41–50 (2000).
- Beleut, M. *et al.* Two distinct mechanisms underlie progesterone-induced proliferation in the mammary gland. *Proc. Natl Acad. Sci. USA* **107**, 2989–2994 (2010).
- Fernandez-Valdivia, R. *et al.* The RANKL signaling axis is sufficient to elicit ductal side-branching and alveologenesis in the mammary gland of the virgin mouse. *Dev. Biol.* **328**, 127–139 (2009).
- Gonzalez-Suarez, E. *et al.* RANK overexpression in transgenic mice with mouse mammary tumor virus promoter-controlled RANK increases proliferation and impairs alveolar differentiation in the mammary epithelia and disrupts lumen formation in cultured epithelial acini. *Mol. Cell. Biol.* **27**, 1442–1454 (2007).
- Aldaz, C. M., Liao, Q. Y., LaBate, M. & Johnston, D. A. Medroxyprogesterone acetate accelerates the development and increases the incidence of mouse mammary tumors induced by dimethylbenzanthracene. *Carcinogenesis* **17**, 2069–2072 (1996).
- Landis, M. W., Pawlyk, B. S., Li, T., Sicinski, P. & Hinds, P. W. Cyclin D1-dependent kinase activity in murine development and mammary tumorigenesis. *Cancer Cell* **9**, 13–22 (2006).
- Wang, T. C. *et al.* Mammary hyperplasia and carcinoma in MMTV-cyclin D1 transgenic mice. *Nature* **369**, 669–671 (1994).
- Cao, Y. *et al.* IKK α provides an essential link between RANK signaling and cyclin D1 expression during mammary gland development. *Cell* **107**, 763–775 (2001).
- Guy, C. T. *et al.* Expression of the *neu* protooncogene in the mammary epithelium of transgenic mice induces metastatic disease. *Proc. Natl Acad. Sci. USA* **89**, 10578–10582 (1992).
- Luo, J. L. *et al.* Nuclear cytokine-activated IKK α controls prostate cancer metastasis by repressing Maspin. *Nature* **446**, 690–694 (2007).
- Herschkowitz, J. I. *et al.* Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol.* **8**, R76 (2007).
- Hewitt, S. C. *et al.* Lack of ductal development in the absence of functional estrogen receptor α delays mammary tumor formation induced by transgenic expression of ErbB2/*neu*. *Cancer Res.* **62**, 2798–2805 (2002).
- Joshi, P. A. *et al.* Progesterone induces adult mammary stem cell expansion. *Nature* **465**, 803–807 (2010).
- Klijn, J. G., Setyono-Han, B. & Foekens, J. A. Progesterone antagonists and progesterone receptor modulators in the treatment of breast cancer. *Steroids* **65**, 825–830 (2000).
- Ismail, P. M. *et al.* Progesterone involvement in breast development and tumorigenesis—as revealed by progesterone receptor “knockout” and “knockin” mouse models. *Steroids* **68**, 779–787 (2003).
- Asselin-Labat, M. L. *et al.* Control of mammary stem cell function by steroid hormone signalling. *Nature* **465**, 798–802 (2010).
- Aupperlee, M., Kariagina, A., Osuch, J. & Haslam, S. Z. Progestins and breast cancer. *Breast Dis.* **24**, 37–57 (2005).
- Hofseth, L. J. *et al.* Hormone replacement therapy with estrogen or estrogen plus medroxyprogesterone acetate is associated with increased epithelial proliferation in the normal postmenopausal breast. *J. Clin. Endocrinol. Metab.* **84**, 4559–4565 (1999).

21. Greendale, G. A. *et al.* Postmenopausal hormone therapy and change in mammographic density. *J. Natl Cancer Inst.* **95**, 30–37 (2003).
22. Chlebowski, R. T. *et al.* Influence of estrogen plus progestin on breast cancer and mammography in healthy postmenopausal women: the Women's Health Initiative Randomized Trial. *J. Am. Med. Assoc.* **289**, 3243–3253 (2003).
23. Holmberg, L. *et al.* Increased risk of recurrence after hormone replacement therapy in breast cancer survivors. *J. Natl Cancer Inst.* **100**, 475–482 (2008).
24. Van Poznak, C. *et al.* Expression of osteoprotegerin (OPG), TNF related apoptosis inducing ligand (TRAIL), and receptor activator of nuclear factor κ B ligand (RANKL) in human breast tumours. *J. Clin. Pathol.* **59**, 56–63 (2006).
25. Briskin, C. *et al.* A paracrine role for the epithelial progesterone receptor in mammary gland development. *Proc. Natl Acad. Sci. USA* **95**, 5076–5081 (1998).
26. Cardiff, R. D. *et al.* The mammary pathology of genetically engineered mice: the consensus report and recommendations from the Annapolis meeting. *Oncogene* **19**, 968–988 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Begley, C. Queva and L. Kirsch for critical reading of the manuscript. The authors would like to thank W. Weng for statistical analysis; H. Li-Ya, L. Woody, K. Stocking, B. Saxbe, D. Hill, L. Cherepow, S. Allred, A. Winters and R. Soriano for technical assistance; M. Blake and M. Tometsko for critical discussions and technical assistance; I. Sarosi, C. Johnson and J. Hawkins for providing tissue specimens; and T. Chang, A. Foreman-Wykert and G. Smith for editorial support.

Author Contributions E.G.-S. and A.P.J. conducted the largest portion of experiments, analysed data, designed experiments and helped to write the manuscript; J.J., R.M., R.E. and M.P.R.-M. conducted experiments and analysed data; J.P., D.B. and W.C.D. designed and supervised the study, analysed data and wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to W.C.D. (dougallw@amgen.com).

METHODS

Mouse experimental procedure. All animal procedures were approved by and performed under the guidelines of Amgen's Institutional Animal Care and Use Committee (IACUC). MMTV-RANK lines 1 and 2 (C57BL/6 background) were established and maintained at Amgen under specific pathogen-free conditions as previously described⁶. Multiparous females were observed for the development of spontaneous tumours up to 29 months of age (>5 pregnancies for each animal). Induction of mammary tumours by MPA and DMBA were in accordance with the published protocol⁷ in which 5–6-week-old female mice were implanted subcutaneously with a 90-day-release 50 mg MPA pellet (Innovative Research of America) and replaced at expiration. One milligram of DMBA (Sigma-Aldrich) was administered at weeks 9, 10, 12, 13 of age by oral gavage. Tumours were detected by manual palpation. For treatment studies, MMTV-RANK line 1 and C57BL/6 wild-type mice were treated with 10 mg kg⁻¹ muRANK-muFc (Amgen Inc.) or PBS subcutaneously three times per week or zoledronic acid (Zometa) dosed at 0.5, 0.2, or 0.025 mg kg⁻¹ weekly. Tumours were detected by manual palpation weekly and latency was expressed as weeks after last DMBA treatment. Mice were killed when a palpable mass exceeded 100 mm² or the third MPA pellet expired. For initial tumorigenesis studies, wild-type and MMTV-RANK lines 1 and 2 mice were killed when a palpable mass exceeded >200 mm² or when mice appeared moribund. MMTV-*neu* (N202 Mul; FvB background) mice were obtained from Jackson Laboratories. The unactivated *neu* proto-oncogene, under control of the MMTV promoter, results in mammary tumours of highly uniform morphology and lung metastasis after a long latency¹¹. Ten milligrams per kilogram RANK-Fc or muFc (IgG1, both Amgen Inc.) were given three times weekly beginning at 5 months of age and continued until mice were killed. Tumours were detected by manual palpation weekly. All mice were injected intraperitoneally with 1 mg per 100 µl BrdU (Sigma) 2 h before euthanasia. For progesterone treatment experiments, mice were given subcutaneous progesterone (1 mg) in sesame oil daily. Mifepristone (10 mg kg⁻¹ in sesame oil subcutaneously) was given daily at least 2 h preceding progesterone treatment. RANK-Fc was given at 10 mg kg⁻¹ subcutaneously every 48 h; the first dose of RANK-Fc was administered 24 h before progesterone. Animals were killed and analysed for BrdU and RANKL immunohistochemistry 72 h after the first progesterone treatment. Demonstration that the 10 mg kg⁻¹ dose of RANK-Fc optimally blocked MPA-induced mammary proliferation was confirmed by a dose-response comparison of 1, 3, 10, 30 and 100 mg kg⁻¹ RANK-Fc *in vivo* (Supplementary Fig. 10). Treatment was initiated with the first DMBA treatment at 9 weeks of age and continued until mice were killed. Previous studies have established the specificity of RANK-Fc for RANKL *in vitro*, via multiple independent direct ligand-binding assays and *in vivo*, using either transgenic overexpression or pharmacological treatment with RANK-Fc^{27,28}.

Whole-mount analysis, histology, immunohistochemistry and mRNA analysis of mammary tissue. Whole mounts of inguinal mammary glands were used to analyse gross morphological changes at the specified times (4 or 7 weeks after the last DMBA treatment) after fixation in Clarke's solution and stained with carmine²⁵. Dense epithelial foci of greater than 300 µm were counted. Haematoxylin and eosin stained sections (4 µm) of inguinal mammary glands from MMTV-RANK, wild-type, or MMTV-*neu* tumour-bearing mice were used to analyse for the presence of pre-neoplastic lesions and overall mammary epithelial density. Tumour histotype was assessed from haematoxylin and eosin stained tissue according to the Annapolis guidelines²⁶. For immunohistochemistry, antigen heat retrieval with DIVA (Biocare Medical) was used on paraffin-embedded sections and sections were incubated with antibodies against muRANKL (AF462; R&D Systems), PR (Dako), CK14 (Covance), CK5 (Abcam 32118), CK18 (Chemicon B3457), cyclin D1 (Lab Vision clone SP4), cleaved caspase 3 (Cell Signaling Technologies catalogue no. 9661) or BrdU (Accurate). For RANK staining, tissues were fixed in Zinc-Tris solution, embedded in paraffin and antigen heat retrieval with DIVA (Biocare Medical) was used. Sections were incubated with muRANK antibody (AF692; R&D Systems). Antigen-antibody complexes were detected with streptavidin horseradish peroxidase (Vector Laboratories) and visualized with diaminobenzidine (Dako). Sections were counterstained with haematoxylin. Dual RANKL and PR stained sections were visualized with alkaline phosphatase Ferangi blue (Biocare Medical) for RANKL and permanent red (DAKO) for PR. The specificity of anti-muRANK and anti-muRANKL antibodies was confirmed by correlation of the immunohistochemistry signal with mRNA expression⁶ and, additionally for anti-RANKL, by demonstrating the absence of staining in skeletal tissues obtained from RANKL-deficient mice. For quantification of BrdU, active-caspase 3, cyclin D1, RANK, RANKL and PR immunostaining, slides were scanned at ×20 using an Aperio ScanScope (Aperio). Algorithms within Spectrum software (Aperio) were used at the default settings to score staining (minimum 500 cells per analysis). MPA-plus-DMBA-treated mammary tissues stained for cyclin D1 were analysed for the percentage (±s.e.m.) of elevated cyclin-D1-positive cells (2+ score) and quantified using Aperio Spectrum software. Total per cent cyclin-D1-positive cells

was quantified in Aperio in MMTV-*neu* treated mice. Expression of survivin (Birc5), RANK and RANKL mRNA was determined via reverse transcriptase polymerase chain reaction (RT-PCR) and primer probe sets from Applied Biosystems.

Tumour transplant experiments. 8-week-old female C57BL/6 mice were implanted with freshly prepared tumour fragments as follows. On experimental day 0, a donor mouse bearing a MMTV-RANK tumour that arose from a tumour fragment implantation was harvested and diced into approximately 2 mm × 2 mm fragments. An incision was made under ketamine/xylazine anaesthesia and a fragment was implanted into a pocket of the mammary fat pad. Mice were also implanted with a 90-day-release MPA pellet (50 mg per pellet, Innovative Research of America). Treatment began on day 0 and consisted of 10 mg kg⁻¹ RANK-Fc or PBS three times per week subcutaneously and continued until termination of the study at day 53. All tumours from this study were harvested at day 53 into neutral buffered formalin for pathology analysis.

Cyclin D1 analysis of primary MECs from mammary glands. Cultured MECs from normal mammary glands were prepared as described⁶. RNA was harvested from MMTV-RANK MECs grown in three-dimensional cultures in differentiation media with or without RANKL for 24 h. Gene expression analysis of cyclin D1 was performed as described⁶. Cyclin D1 expression was assessed by immunohistochemistry 72 h after RANKL treatment.

Immunofluorescence analysis of primary MECs from mammary glands. Mammary glands were dissected from day 16.5 pregnant mice as described⁶. Briefly, tissues were minced in DMEM/F12 (Gibco) containing antibiotics, and incubated with 0.3% collagenase (Roche), and 2.5 U ml⁻¹ dispase (Invitrogen) at 37 °C for 90 min with trituration every 30 min. Crude digests were filtered through 500 AM Nitex mesh (Technicon) to remove undigested clumps, and centrifuged at 100g to collect epithelial spheroids. Pellets were washed five times in DMEM/F12 with antibiotics and 5% fetal bovine serum (BCS) and collected each time with a low-speed spin. Spheroids were plated on 10-cm plates in growth medium (DMEM/F12 with antibiotics, 10 µg ml⁻¹ insulin, 10 ng ml⁻¹ EGF, and 5% BCS), for 2 to 5 days before being harvested for immunofluorescent staining (Cellomics) or for three-dimensional culture. The culture medium was changed every other day. For three-dimensional cultures, first passage primary mammary epithelial cells (MECs) were harvested by trypsinization. Cells were counted and replated in MatrigelR (Collaborative Research) coated onto 8-well chamber slides (LabTek) and growth medium was added. After 24 h, the medium was changed to differentiation medium (DMEM/F12 with antibiotics, liquid media supplement ITS (Sigma), 3 µg ml⁻¹ prolactin (Sigma), 1 µg ml⁻¹ hydrocortisone (Sigma) with or without murine RANKL-LZ (200 ng ml⁻¹); medium was changed every day.

Acinar structures were stained as previously described⁶. MEC organoids were cultured in Matrigel in 8-well chamber slides. At day 8 the medium was removed and acini were fixed in 2% paraformaldehyde and permeabilized using 0.5% Triton X-100 before blocking. Cells were incubated with the primary antibodies rat anti-integrin α6 (1:150; Chemicon MAB1378) and rabbit anti ZO-1 (1:100; Zymed, 61-7300) overnight at room temperature and then with Alexa-conjugated secondary antibodies (Molecular Probes) for 45 min at room temperature and DRAQ5 (Alexis) for nuclear staining. Slides were mounted with Prolong Gold Antifade Reagent (Molecular Probes). Confocal analysis was performed using the Zeiss confocal microscopy system equipped with argon and HeNe lasers. Images were captured using LSM TM version 5 software (Zeiss).

Immunohistochemistry of human breast cancer samples. Anti-human RANK and RANKL immunohistochemistry was performed on sections from formalin-fixed, paraffin-embedded tumour specimens (57 samples of invasive carcinoma of human breast: 39 ductal carcinoma, 6 lobular, 3 mucinous, 9 non-specified) prepared with heat retrieval in Diva buffer or citrate buffer pH 6 for 20 min. Immunohistochemistry was performed with an automated staining method using Lab Vision or Dako Autostainers. The primary antibody concentration was 0.75 µg ml⁻¹ for anti-huRANKL monoclonal antibody (M366; Amgen) and 5 µg ml⁻¹ for anti-human RANK (N-1H8 and N-2B10) with goat anti-mouse secondary antibody. An isotype control mouse IgG1 and appropriate secondary antibody was performed for each specimen and each antibody. Antibody reactivity was visualized by diaminobenzidine enhanced by tyramide amplification. A formalin-fixed giant cell tumour of the bone and transfected cell lines were used as positive controls for the anti-RANKL and anti-RANK antibodies. Specificity of the anti-human RANKL and anti-human RANK monoclonal antibodies were further substantiated by flow cytometry and western blot analysis of positive and negative control cells. The low frequency of human tumours with RANKL-positive tumour epithelia precluded any correlative analysis with PR expression. RANKL-positive fibroblast-like stromal cells were only observed within a dense lymphoid environment. Antibody staining for RANK and RANKL was scored on a scale of 0 to 3 for intensity (0 = no staining, 1 = weak signal, 2 = moderate stain, 3 = intense stain) and for the percentage of positive cells. A complex score was used to analyse the epithelial component of tumours and was equal to the percentage of positive cells times the

intensity score. A tumour was considered positive above a threshold of 10. For RANKL only the most intensely staining cells (=3) were counted.

Statistical analyses. The log-rank test was used to assess differences in kinetics of tumour formation. Fisher's exact test was used for categorical analysis of differences in tumour phenotype. One-way ANOVA with Bonferroni's post-test was used to evaluate differences in immunohistochemistry-quantified sections and to evaluate differences in other multiple group comparisons. For all other numerical

comparisons, independent *t*-tests were used. For all tests, $P < 0.05$ was considered significant.

27. Anderson, D. M. *et al.* A homologue of the TNF receptor and its ligand enhance T-cell growth and dendritic-cell function. *Nature* **390**, 175–179 (1997).
28. Hsu, H. *et al.* Tumor necrosis factor receptor family member RANK mediates osteoclast differentiation and activation induced by osteoprotegerin ligand. *Proc. Natl Acad. Sci. USA* **96**, 3540–3545 (1999).

A methyl transferase links the circadian clock to the regulation of alternative splicing

Sabrina E. Sanchez^{1*}, Ezequiel Petrillo^{2*}, Esteban J. Beckwith³, Xu Zhang⁴, Matias L. Rugnone¹, C. Esteban Hernando¹, Juan C. Cuevas⁵, Micaela A. Godoy Herz², Ana Depetris-Chauvin³, Craig G. Simpson⁶, John W. S. Brown^{6,7}, Pablo D. Cerdán³, Justin O. Borevitz⁴, Paloma Mas⁵, M. Fernanda Ceriani³, Alberto R. Kornblihtt² & Marcelo J. Yanovsky^{1,3}

Circadian rhythms allow organisms to time biological processes to the most appropriate phases of the day–night cycle¹. Post-transcriptional regulation is emerging as an important component of circadian networks^{2–6}, but the molecular mechanisms linking the circadian clock to the control of RNA processing are largely unknown. Here we show that PROTEIN ARGININE METHYL TRANSFERASE 5 (PRMT5), which transfers methyl groups to arginine residues present in histones⁷ and Sm spliceosomal proteins^{8,9}, links the circadian clock to the control of alternative splicing in plants. Mutations in *PRMT5* impair several circadian rhythms in *Arabidopsis thaliana* and this phenotype is caused, at least in part, by a strong alteration in alternative splicing of the core-clock gene *PSEUDO RESPONSE REGULATOR 9* (*PRR9*). Furthermore, genome-wide studies show that PRMT5 contributes to the regulation of many pre-messenger-RNA splicing events, probably by modulating 5′-splice-site recognition. *PRMT5* expression shows daily and circadian oscillations, and this contributes to the mediation of the circadian regulation of expression and alternative splicing of a subset of genes. Circadian rhythms in locomotor activity are also disrupted in *dart5-1*, a mutant affected in the *Drosophila melanogaster* PRMT5 homologue, and this is associated with alterations in splicing of the core-clock gene *period* and several clock-associated genes. Our results demonstrate a key role for PRMT5 in the regulation of alternative splicing and indicate that the interplay between the circadian clock and the regulation of alternative splicing by PRMT5 constitutes a common mechanism that helps organisms to synchronize physiological processes with daily changes in environmental conditions.

The circadian clock interacts with light to inhibit stem elongation in plants¹⁰, so we used tall plants under conditions of a light–dark cycle to identify novel clock mutants. This screening, combined with a positional cloning approach, allowed us to find a mutation in *PRMT5* that lengthened the period of circadian rhythms in leaf movement and gene expression in *Arabidopsis* (Fig. 1a, b and Supplementary Fig. 1). The same phenotype was observed in other loss-of-function alleles of *PRMT5* (Supplementary Fig. 1).

The core oscillator in *Arabidopsis* depends on the interaction between the Myb transcription factors CIRCADIAN CLOCK ASSOCIATED 1 (*CCA1*) and LATE ELONGATED HYPOCOTYL (*LHY*), and the PSEUDO-RESPONSE REGULATOR 1 (*PRR1*) (also known as TIMING OF CAB EXPRESSION 1 (*TOC1*))¹¹. We observed that *prmt5* mutants had a longer circadian period of *TOC1* expression than wild-type plants (Fig. 1c and Supplementary Fig. 2), and a delayed expression of *TOC1* in a light–dark cycle (Supplementary Fig. 2). We observed similar alterations in *LHY* and *CCA1* expression (Supplementary Fig. 3). Interestingly, *PRMT5* expression was regulated by the circadian clock (Fig. 1d and Supplementary Fig. 4), indicating that it is part of a feedback loop controlling clock function in *Arabidopsis*.

PRMT5 methylates histone 4 in *Arabidopsis* and regulates flowering time through the epigenetic repression of *FLOWERING LOCUS C* (*FLC*)^{12–14}. *FLC* regulates circadian rhythms at high temperatures¹⁵, but genetic analysis indicated that *prmt5* affects circadian rhythms independently of *FLC* (Supplementary Fig. 5). Next we explored the connection between PRMT5 and the circadian network more thoroughly using Affymetrix expression arrays. Clock-regulated genes¹⁵ were over-represented among PRMT5 targets at the mRNA level (1.7-fold enrichment; $P < 3 \times 10^{-11}$). Indeed, genes upregulated in *prmt5* mutants were enriched in those with peak expression during daytime, and genes downregulated were enriched in those with peak expression during night time (Supplementary Fig. 6). More interestingly, the gene showing the largest enhancement in expression was *PRR9* (Supplementary Fig. 7a and Supplementary Table 1), a homologue of *TOC1/PRR1* that acts redundantly with *PRR7* to repress *CCA1* and *LHY* expression¹⁶. *PRR7* mRNA levels were also increased in *prmt5-5* mutants (Supplementary Table 2). Quantitative polymerase chain reaction (qPCR) analysis confirmed increased *PRR9* mRNA levels in *prmt5-5* plants, particularly in the late subjective day (Supplementary Fig. 7b). Overexpression of *PRR9*, however, shortens rather

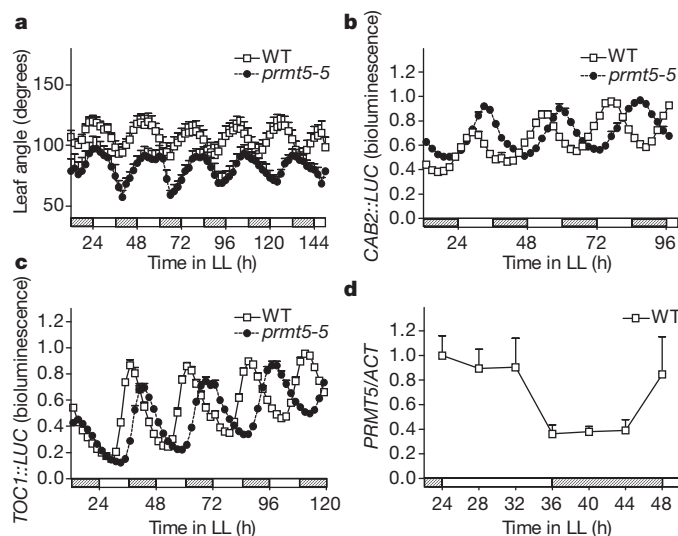


Figure 1 | A role for PRMT5 in the circadian system of *Arabidopsis*. **a**, Circadian rhythm of leaf movement in constant light (LL) (h, hours; $n = 6$). WT, wild type. **b**, **c**, Bioluminescence analysis of *CAB2::LUC* expression ($n = 7–10$) (**b**) or *TOC1::LUC* expression ($n = 8–10$) (**c**). **d**, *PRMT5* expression measured by qPCR in constant light ($n = 4$) relative to *ACTIN 8* (*ACT*). Data represent average \pm s.e.m. Open squares, wild type; filled circles, *prmt5* mutant. Open and lined boxes indicate subjective day or night, respectively.

¹IFEVA, Facultad de Agronomía, UBA-CONICET, C1417DSE Buenos Aires, Argentina. ²FIBYNE, FCEyN, UBA-CONICET, C1428EGA Buenos Aires, Argentina. ³Fundación Instituto Leloir, IIBBA-CONICET, C1405BWE Buenos Aires, Argentina. ⁴Department of Ecology & Evolution, University of Chicago, Chicago, Illinois 60637, USA. ⁵Centre for Research in Agricultural Genomics (CRAG), Consortium CSIC-IRTA-UAB, Barcelona 08034, Spain. ⁶Genetics Programme, SCRI, Invergowrie, Dundee DD2 5DA, UK. ⁷Division of Plant Sciences, University of Dundee at SCRI, Dundee, DD2 5DA, UK.

*These authors contributed equally to this work.

than lengthens the period of circadian rhythms¹⁷. Furthermore, we could not detect a PRMT5 epigenetic mark⁷ in *PRR9* (Supplementary Fig. 8), indicating that circadian alterations in *prmt5* do not result from defective epigenetic repression of *PRR9*.

In addition to histones, PRMT5 (also known as DART5/CSUL) symmetrically dimethylates Sm proteins in *Drosophila*^{8,9} and we obtained similar evidence in *Arabidopsis* (Supplementary Fig. 9). Indeed, PRMT5 acts together with the survival motor neuron (SMN) complex to regulate spliceosome assembly in human cells¹⁸, and mutations in SMN have been shown to affect alternative splicing in a tissue-specific manner in mice¹⁹. However, the lack of symmetrical dimethylation of Sm proteins in *prmt5* mutant flies does not impair small nuclear ribonucleoprotein particle (snRNP) assembly²⁰. Therefore, if and how PRMT5 contributes to the regulation of pre-mRNA splicing is uncertain.

Besides complementary DNAs encoding the full-length *PRR9* protein, splice variants that retain intron 3 and/or use an alternative 5' splice site at the end of exon 2 have been reported (Fig. 2a). A PCR preceded by a reverse transcription reaction (RT-PCR) conducted to evaluate the relative abundance of all *PRR9* mRNA isoforms revealed significant alterations in alternative splicing in *prmt5* mutants (Fig. 2a and Supplementary Fig. 10). qPCR confirmed that *prmt5* seedlings had increased levels of an isoform that retains intron 3 (Fig. 2b), but greatly reduced levels of the isoform encoding the full-length protein (Fig. 2c). No effect on circadian rhythmicity was observed in plants overexpressing *PRR9_B* mRNA, which encodes a truncated protein identical to that of *PRR9_D* mRNA (Supplementary Fig. 11). On the other hand, reduced levels of the

functional *PRR9* mRNA isoform may contribute to the phenotype of the mutant, as *prmt5-5* delayed the pace of the clock on a wild-type background but did not affect circadian period on a *prr9;prr7* background, where the function of *PRR9* is most evident owing to its redundancy with *PRR7*¹⁶ (Fig. 2d). Interestingly, *PRR9_A* and *PRR9_B* mRNAs, the most abundant isoforms in wild-type plants, oscillate with slightly different phases (Supplementary Fig. 12), indicating that this may contribute to refining the shape of *PRR9* oscillation, ensuring that repression of *CCA1* and *LHY* expression begins after dawn. Changes in alternative splicing of *PRR9*, however, are unlikely to account for all the effect of *prmt5* on circadian rhythms, as both *PRR9* and *PRR7* are required for PRMT5 effects on the clock (Fig. 2d) and PRMT5 affects *PRR7* expression (Supplementary Table 2 and Supplementary Fig. 13) but not its splicing (data not shown). Taken together, these results indicate that PRMT5 affects circadian rhythms in *Arabidopsis*, regulating *PRR9* alternative splicing as well as the timing of expression and/or function of *PRR7*.

Next we analysed the effects of PRMT5 on pre-mRNA splicing at a genome-wide level using Affymetrix tiling arrays. We identified 471 introns, out of a total of 67,791, with significantly higher hybridization signals in *prmt5* mutants (Supplementary Table 3). The increased signal for these introns probably reflects intron retention, as most of the other introns and exons within these genes did not differ between wild-type and mutant plants (Fig. 3a). These results were confirmed for a subset of genes by conventional RT-PCR (Supplementary Fig. 14). To determine changes in alternative splicing in more detail we used a high-resolution RT-PCR panel that included several known alternative splicing events^{21,22}. *prmt5* mutants had significant alterations in 44 out of 288 events analysed (Supplementary Table 4), which included exon skipping, alternative donor and acceptor splice sites, as well as intron retention (Fig. 3b and Supplementary Table 4). Interestingly, *PRR9* was one of the genes showing the largest alterations in alternative splicing in *prmt5* mutants (Supplementary Table 4 and Supplementary Fig. 15). The fact that PRMT5 regulated approximately 15% of the alternative splicing events analysed through the high-resolution RT-PCR panels, but only a minor proportion of introns evaluated using tiling arrays (less than 1%), indicates that this protein has specific functions regulating splice-site selection rather than a role in constitutive pre-mRNA splicing. Interestingly, we found a significant enrichment in alternative 5' splice sites among all splicing events affected in *prmt5* mutants (Fig. 3c). A defect in 5'-splice-site recognition should also cause retention of some introns and, indeed, all the splicing defects detected using tiling arrays corresponded to increased intron retention (Supplementary Table 3). A comparison between the 5' splice site of the splicing events most significantly affected in *prmt5* mutants (Supplementary Table 5) and the *Arabidopsis* 5'-splice-site consensus sequence, revealed an important decrease in the frequency of the dominant G at the -1 position and a tendency towards randomization of the nucleotides present at the -2, -1 and +5 positions (Fig. 3d). Thus, these observations strongly indicate that PRMT5 regulates alternative splicing, contributing, at least in part, to the recognition of weak 5' splice sites.

Given that the circadian clock regulates *PRMT5* expression and that PRMT5 contributes to modulate alternative splicing, we proposed that PRMT5 may link the circadian clock to the regulation of alternative splicing. To evaluate this possibility in detail we analysed alternative splicing of *RUBISCO ACTIVASE* (*RCA*) under light-dark-cycle and free-running conditions. *RCA* produces a short mRNA encoding a protein whose activity is regulated by light intensity and a long mRNA that arises through the usage of an alternative donor site, encoding a protein that is not regulated by light²³. Our RT-PCR analysis confirmed the alterations in *RCA* alternative splicing observed in *prmt5-5* using the high-resolution RT-PCR panel (Fig. 3e). Furthermore, we found that the ratio between the mRNA levels of the two *RCA* splice variants showed daily and circadian oscillations in wild-type *Arabidopsis*. These oscillations seem to be physiologically relevant, as the relative proportion of the short mRNA, encoding the protein whose activity is regulated by light intensity, increased during daytime (Fig. 3f).

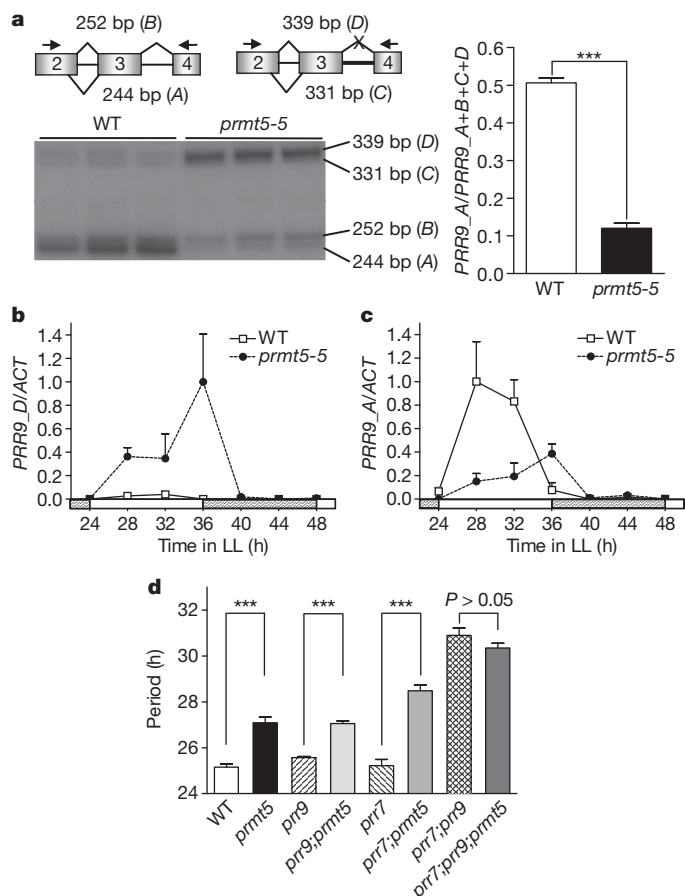


Figure 2 | PRMT5 affects expression and alternative splicing of the clock gene *PRR9*. **a**, *PRR9* splice variants (*PRR9_A* to *PRR9_D*) and their relative abundance measured by radioactive PCR (***) $P < 0.0001$; $n = 3$). bp, base pairs. **b**, **c**, Expression of *PRR9_D* (**b**) and *PRR9_A* (**c**) by qPCR in constant light ($n = 4$). **d**, Free-running period of leaf movement (***) $P < 0.001$; $n = 6$). Data represents average + s.e.m. Open squares, wild type; filled circles, *prmt5* mutant.

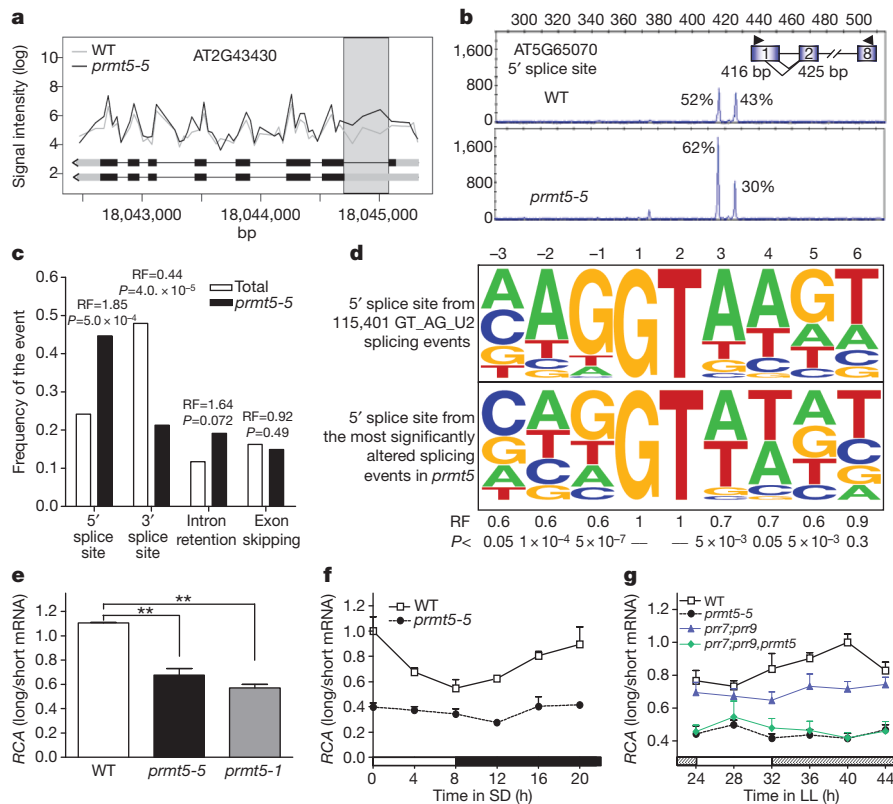


Figure 3 | Genome-wide analysis of pre-mRNA splicing in *Arabidopsis prmt5* mutants. **a**, Tiling array data of a representative gene with an alteration in pre-mRNA splicing in *prmt5* mutants ($n = 3$). **b**, Example of a high-resolution RT-PCR reaction shown as an electropherogram depicting splicing pattern changes in *prmt5-5*. Percentage of each isoform relative to the total is shown. **c**, Frequency of each type of alternative splicing event in the high-resolution RT-PCR panel (total) and among those altered in *prmt5-5*. The representation factor (RF) is the frequency in *prmt5-5* divided by the total frequency. **d**, Pictograms showing the frequency distribution of nucleotides at

the 5' splice site of 115,401 *Arabidopsis* introns (top), and in the 50 splicing events most significantly altered in *prmt5-5* (bottom). **e–g**, RCA splice variants in plants grown always in constant light (** $P < 0.01$) (**e**), in a short-day cycle (SD; 8:16 h light–dark) (**f**) or in seedlings transferred to constant light after entrainment in a light–dark cycle (**g**). Values are average \pm s.e.m. ($n = 3–4$). Open squares, wild type; filled circles, *prmt5-5*; uptriangles, *prr7;prr9*; diamonds, *prr7;prr9;prmt5*. Open and closed boxes indicate light and dark period, respectively. Lined boxes indicate subjective night.

Circadian oscillations in RCA alternative splicing were disrupted in *prmt5* seedlings and, more importantly, alterations in the ratio of the two mRNA isoforms were more severe in *prmt5* and *prmt5;prr9;prr7* than in *prr9;prr7* mutants, indicating a direct role for PRMT5 in the control of this process rather than an indirect role through its effect on circadian rhythms (Fig. 3g). Circadian oscillations in alternative splicing were also observed for *RSP31*, but not for *AT5G57630*, two alternatively spliced genes regulated by PRMT5 according to high-resolution RT-PCR analysis (Supplementary Fig. 16). Lack of oscillation of some PRMT5 targets may be due to enhanced stability of the corresponding mRNAs.

Two mutant alleles affected in the *D. melanogaster prmt5* gene, *dart5-1* and *csul^{RM50}*, have been identified^{8,9}, and allowed us to evaluate the effects of PRMT5 on the circadian system and splicing regulation in this organism. Compared to heterozygous controls, homozygous *dart5* mutants showed poor rhythms in locomotor activity in a light–dark cycle and decreased rhythmicity in constant darkness (Fig. 4a, b). Similar alterations were observed for *csul^{RM50}* (Supplementary Fig. 17). However, we observed strong oscillations in the expression of core-clock genes and in PERIOD (PER) protein levels (Fig. 4c, d and Supplementary Fig. 18), indicating that PRMT5 modulates behavioural rhythms and thus controls the mechanism linking the central oscillator to clock output, rather than through effects on the core oscillator.

Next we evaluated pre-mRNA splicing of clock and non-clock genes in *Drosophila*. Circadian regulation of alternative splicing of the core-clock gene *per* has been reported to have an important role in adjusting circadian rhythms to changes in photoperiod and temperature in *Drosophila*^{5,6}. Interestingly, we observed circadian oscillations in the

ratio of the two *per* mRNA variants⁵ in wild-type flies (that is, intron-8-retained A or a spliced variant B'), whereas mostly B' was present in *dart5-1* flies independently of the time of day (Fig. 4e). We also conducted a genome-wide analysis of pre-mRNA splicing using tiling arrays and found 418 introns whose hybridization signal increased in *dart5-1*, several of which were confirmed by qPCR (Fig. 4f), and 30 introns that showed higher signals in wild-type flies (Supplementary Table 6). Among the genes that showed increased intron retention in *dart5-1* we found *takeout* (*to*) and *slowpoke*, two genes implicated in the regulation of clock outputs in *Drosophila*^{24,25}, as well as *norpA*, a phospholipase C that regulates clock entrainment²⁶ and *per* splicing (Supplementary Table 6)^{5,6,26}. There were also differences in the expression of 65 exons corresponding to genes whose remaining exons, in most cases, did not show differences in their overall levels (Supplementary Table 7). In contrast to *dart5-1*, the arrhythmic mutants *per⁰¹* and *clk^{irk}* did not show large alterations in splicing, at least for the events evaluated here, indicating that changes in splicing in *dart5-1* are not simply the indirect consequence of arrhythmicity (Supplementary Fig. 19).

Similar to the findings in *Arabidopsis*, we observed a reduction in the frequency of the dominant G –1 nucleotide in the 5'-splice-site sequences of the most significantly affected splicing events in *dart5-1* (Fig. 4g and Supplementary Table 8). Taken together, these results strongly indicate that PRMT5 regulates pre-mRNA splicing, affecting the efficiency with which snRNPs interact with certain splice sites. Recognition of 5' splice sites involves direct RNA–RNA interactions between the 5' end of the U1 small nuclear RNA and the 5'-splice-site sequence of the target mRNAs. SmB, Smd1 and Smd3 make direct

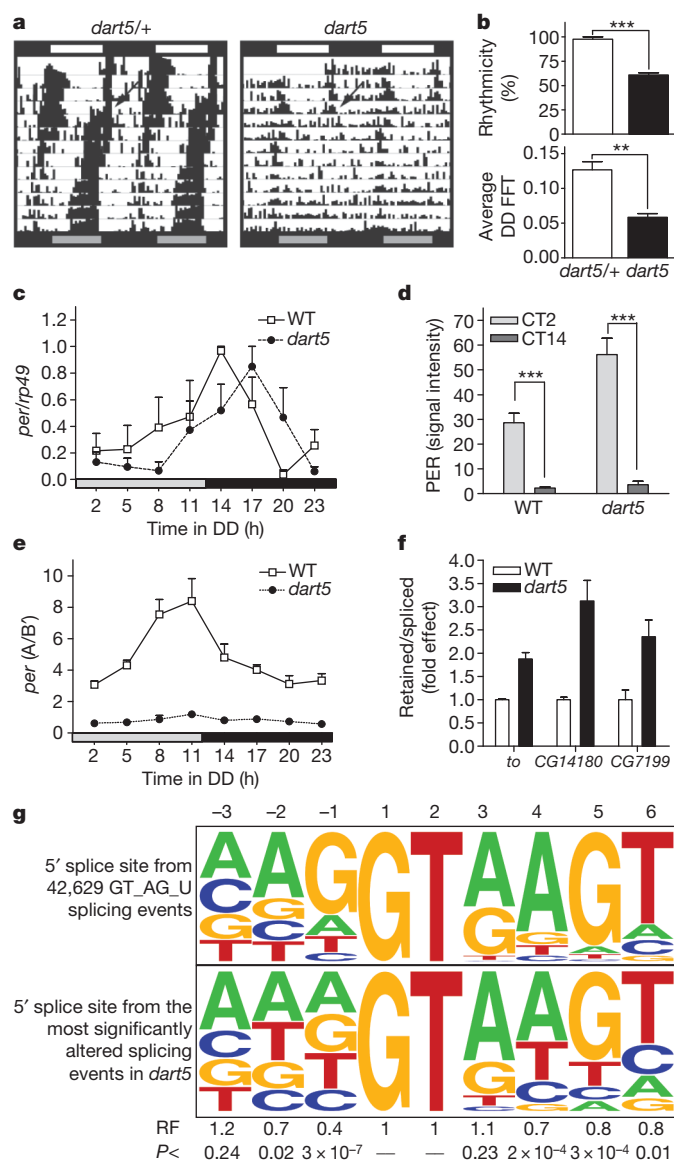


Figure 4 | PRMT5 regulates circadian rhythms and pre-mRNA splicing in *Drosophila*. **a**, Representative double-plotted actograms. Arrows indicate transfer to constant darkness (DD). **b**, Percentage of rhythmic flies (top) and power fast Fourier transform (FFT) (bottom). *** $P < 0.001$, ** $P < 0.01$; $n = 70$. **c**, *per* expression measured by qPCR in constant darkness ($n = 4$). **d**, Nuclear PER immunoreactivity on the small ventral lateral neurons of the *Drosophila* brain in constant darkness. *** $P < 0.0001$ ($n \geq 17$). CT, circadian time. **e**, Relative levels of A and B' *per* mRNA isoforms over a circadian cycle in constant darkness ($n = 4$). **f**, Radioactive PCR validation of intron retention events for several genes identified using tiling arrays. **g**, Pictograms showing the frequency distribution of nucleotides at the 5' splice site of 42,629 introns in wild-type *Drosophila* genes (top) and in the 50 splicing events most significantly altered in *dart5* (bottom). RF is the frequency observed in *dart5* affected events divided by the genome-wide frequency. **b–f**, Data are average \pm s.e.m. Open squares, wild type (w^{1118}); filled circles, *dart5* (*dart5-1*). **a**, **c**, **e**, black and white boxes represent night and day phases, respectively. Grey bars, subjective day.

contact with the pre-mRNA substrate, close to the 5' splice site and act by stabilizing RNA–RNA interactions in yeast²⁷. Thus, methylation of Sm proteins by PRMT5 in plants and flies may facilitate recognition of a subset of 5' splice sites, stabilizing weak RNA–RNA interactions between the U1 snRNA and the 5' splice site of target mRNA. However, we cannot exclude the possibility that some alterations in alternative splicing result from direct or indirect epigenetic defects in *prmt5* mutants²⁸.

We also evaluated global effects of PRMT5 on gene expression in *Drosophila* and found 315 genes downregulated and 85 genes upregulated in *dart5-1* (Supplementary Table 9), including a ninefold increase in the mRNA levels of *to* (Supplementary Fig. 20). In contrast to *Arabidopsis*, we did not observe an over-representation of circadianly regulated genes among PRMT5 target genes in *Drosophila* (Supplementary Fig. 21). This was consistent with the lack of circadian regulation of *prmt5/dart5* expression in *Drosophila* (Supplementary Fig. 22). Nonetheless, we did observe a strong enrichment in genes regulated by the core oscillator component CLOCK among PRMT5-regulated genes²⁹ (Supplementary Fig. 23 and Supplementary Table 10). Thus, although PRMT5 is less tightly linked to the circadian system in *Drosophila* compared to what we observed for PRMT5 in *Arabidopsis*, it is still deeply associated with the fly circadian network controlling behavioural rhythms, splicing of clock-associated genes, and expression of a subset of output genes regulated by the protein CLOCK.

Despite general similarities in the molecular architecture of circadian networks across distantly related organisms, a lack of homology between most clock components indicates an independent evolutionary origin of circadian systems. The existence of a few genes with common roles modulating circadian rhythms in plants and animals³⁰ probably reflects convergent evolutionary processes associated with a limited number of proteins capable of adapting complex signalling networks to common environmental challenges. The dual role of PRMT5 as an epigenetic regulator of gene expression and as a regulator of alternative splicing makes this protein an ideal tool to link a diverse set of molecular and physiological processes to a central oscillatory system, helping organisms to adjust their growth and development to daily changes in their environment.

METHODS SUMMARY

Plant material and physiological assays. All the *Arabidopsis* lines used in this study were Columbia ecotype. Seedlings were grown at a constant temperature of 22 °C under different light regimes depending on the experiment. Detailed information about flowering time, leaf movement and bioluminescence assays is found in Methods.

***Drosophila* behavioural assays.** Flies were grown and maintained at 25 °C and kept in 12:12 h light–dark cycles. Activity was monitored in a light–dark cycle for 4 days and then in constant darkness for 8 days. Data were collected using activity monitors (TriKinetics). Rhythmicity and fast Fourier transform (FFT) analysis were performed using the ClockLab software (Actimetrics) from data collected in constant darkness.

Genome tiling array analysis. Total RNA was processed and hybridized to GeneChip *Arabidopsis* ATH1 Genome Arrays, GeneChip *Arabidopsis* Tiling 1.0R Array, or GeneChip *Drosophila* Tiling 1.0R Array (Affymetrix) according to the manufacturer's instructions. Detailed information about statistical analysis is found in Methods.

qRT–PCR expression analysis. RQ1 RNase-Free DNase (Promega) treated RNA samples were subjected to retrotranscription using Invitrogen SuperScript III and oligo-dT. Synthesized cDNAs were amplified with FastStart Universal SYBR Green Master (Roche) using the 7500 Real Time PCR System (Applied Biosystems) cyclor. *ACTIN 8* and *ribosomal protein 49* (*rp49*) transcripts were used as house-keeping for *Arabidopsis* and *Drosophila*, respectively.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 18 September 2009; accepted 1 September 2010.

Published online 20 October 2010.

- Young, M. W. & Kay, S. A. Time zones: a comparative genetics of circadian clocks. *Nature Rev. Genet.* **2**, 702–715 (2001).
- Hazen, S. *et al.* Exploring the transcriptional landscape of plant circadian rhythms using genome tiling arrays. *Genome Biol.* **10**, R17 (2009).
- Schoning, J. C., Streitner, C., Meyer, I. M., Gao, Y. & Staiger, D. Reciprocal regulation of glycine-rich RNA-binding proteins via an interlocked feedback loop coupling alternative splicing to nonsense-mediated decay in *Arabidopsis*. *Nucleic Acids Res.* **36**, 6977–6987 (2008).
- Guo, J., Cheng, P., Yuan, H. & Liu, Y. The exosome regulates circadian gene expression in a posttranscriptional negative feedback loop. *Cell* **138**, 1236–1246 (2009).

5. Majercak, J., Chen, W. & Edery, I. Splicing of the *period* gene 3'-terminal intron is regulated by light, circadian clock factors, and phospholipase C. *Mol. Cell. Biol.* **24**, 3359–3372 (2004).
6. Collins, B. H., Rosato, E. & Kyriacou, C. P. Seasonal behavior in *Drosophila melanogaster* requires the photoreceptors, the circadian clock, and phospholipase C. *Proc. Natl Acad. Sci. USA* **101**, 1945–1950 (2004).
7. Bedford, M. & Richard, S. Arginine methylation: an emerging regulator of protein function. *Mol. Cell* **18**, 263–272 (2005).
8. Gonsalvez, G., Rajendra, T., Tian, L. & Matera, A. The Sm-protein methyltransferase, Dart5, is essential for germ-cell specification and maintenance. *Curr. Biol.* **16**, 1077–1089 (2006).
9. Anne, J., Ollio, R., Ephrussi, A. & Mechler, B. Arginine methyltransferase Capsuléen is essential for methylation of spliceosomal Sm proteins and germ cell formation in *Drosophila*. *Development* **134**, 137–146 (2007).
10. Nozue, K. *et al.* Rhythmic growth explained by coincidence between internal and external cues. *Nature* **448**, 358–361 (2007).
11. Alabadi, D. *et al.* Reciprocal regulation between TOC1 and LHY/CCA1 within the *Arabidopsis* circadian clock. *Science* **293**, 880–883 (2001).
12. Pei, Y. *et al.* Mutations in the type II protein arginine methyltransferase AtPRMT5 result in pleiotropic developmental defects in *Arabidopsis*. *Plant Physiol.* **144**, 1913–1923 (2007).
13. Schmitz, R., Sung, S. & Amasino, R. Histone arginine methylation is required for vernalization-induced epigenetic silencing of FLC in winter-annual *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **105**, 411–416 (2008).
14. Wang, X. *et al.* SKB1-mediated symmetric dimethylation of histone H4R3 controls flowering time in *Arabidopsis*. *EMBO J.* **26**, 1934–1941 (2007).
15. Edwards, K. D. *et al.* FLOWERING LOCUS C mediates natural variation in the high-temperature response of the *Arabidopsis* circadian clock. *Plant Cell* **18**, 639–650 (2006).
16. Farré, E. M., Harmer, S. L., Harmon, F. G., Yanovsky, M. J. & Kay, S. A. Overlapping and distinct roles of *PRR7* and *PRR9* in the *Arabidopsis* circadian clock. *Curr. Biol.* **15**, 47–54 (2005).
17. Matsushika, A., Imamura, A., Yamashino, T. & Mizuno, T. Aberrant expression of the light-inducible and circadian-regulated *APRR9* gene belonging to the circadian-associated *APRR1/TOC1* quintet results in the phenotype of early flowering in *Arabidopsis thaliana*. *Plant Cell Physiol.* **43**, 833–843 (2002).
18. Chari, A. *et al.* An assembly chaperone collaborates with the SMN complex to generate spliceosomal SnRNPs. *Cell* **135**, 497–509 (2008).
19. Zhang, Z. *et al.* SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. *Cell* **133**, 585–600 (2008).
20. Gonsalvez, G. B., Praveen, K., Hicks, A. J., Tian, L. & Matera, A. G. Sm protein methylation is dispensable for snRNP assembly in *Drosophila melanogaster*. *RNA* **14**, 878–887 (2008).
21. Simpson, C. G. *et al.* Monitoring changes in alternative precursor messenger RNA splicing in multiple gene transcripts. *Plant J.* **53**, 1035–1048 (2008).
22. Raczyńska, K. D. *et al.* Involvement of the nuclear cap-binding protein complex in alternative splicing in *Arabidopsis thaliana*. *Nucleic Acids Res.* **38**, 265–278 (2010).
23. Zhang, N., Kallis, R. P., Ewy, R. G. & Portis, A. R. Light modulation of Rubisco in *Arabidopsis* requires a capacity for redox regulation of the larger Rubisco activase isoform. *Proc. Natl Acad. Sci. USA* **99**, 3330–3334 (2002).
24. Sarov-Blat, L., So, W., Liu, L. & Rosbash, M. The *Drosophila* takeout gene is a novel molecular link between circadian rhythms and feeding behavior. *Cell* **101**, 647–656 (2000).
25. Ceriani, M. F. *et al.* Genome-wide expression analysis in *Drosophila* reveals genes controlling circadian behavior. *J. Neurosci.* **22**, 9305–9319 (2002).
26. Glaser, F. T. & Stanewsky, R. Temperature synchronization of the *Drosophila* circadian clock. *Curr. Biol.* **15**, 1352–1363 (2005).
27. Zhang, D., Abovich, N. & Rosbash, M. A biochemical function for the Sm complex. *Mol. Cell* **7**, 319–329 (2001).
28. Schor, I. E., Rascovan, N., Pelisch, F., Alló, M. & Kornblihtt, A. R. Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *Proc. Natl Acad. Sci. USA* **106**, 4325–4330 (2009).
29. McDonald, M. & Rosbash, M. Microarray analysis and organization of circadian gene expression in *Drosophila*. *Cell* **107**, 567–578 (2001).
30. Rosbash, M. The implications of multiple circadian clock origins. *PLoS Biol.* **7**, e62 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. M. Farré and S. L. Harmer for seeds; J. J. Casal and S. Mora-García for critical reading of the manuscript; the laboratories of A.R.K. and M.J.Y. for discussions; J. Fuller and P. Tondi for technical assistance. This work was supported by grants from the Fundación Antorchas, the Agencia Nacional de Promoción de Ciencia y Tecnología de Argentina, the Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina and the University of Buenos Aires to M.J.Y. and A.R.K., and from the European Union Network of Excellence on Alternative Splicing (EURASNET) to A.R.K. and J.W.S.B. M.J.Y. and A.R.K. are Howard Hughes Medical Institute international research scholars. P.M.'s laboratory is supported by a Ministerio de Educación y Ciencia grant, the European Young Investigator Awards and the EMBO Young Investigator Awards; M.F.C.'s laboratory is supported by Proyecto de Investigación Científica y Tecnológica 2006-1249.

Author Contributions S.E.S. and E.P. contributed equally to this work and performed most of the experiments in this study, with technical assistance from M.L.R., C.E.H., M.A.G.H. and P.D.C. J.C.C. and P.M. conducted bioluminescence and chromatin immunoprecipitation assays. E.J.B. and A.D.C. performed locomotor activity assays, RNA extractions and immunostaining experiments in *Drosophila*. E.P., C.G.S. and J.W.S.B. performed high-resolution RT-PCR experiments. X.Z. and J.O.B. analysed tiling array data. S.E.S., E.P., M.F.C., A.R.K. and M.J.Y. provided input in the preparation of the manuscript, and S.E.S., E.P. and M.J.Y. wrote the paper.

Author Information The data discussed in this publication have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus and are accessible through GEO Series accession number GSE18808 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE18808>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.J.Y. (yanovsky@agro.uba.ar).

METHODS

Plant material and growth conditions. All the *Arabidopsis* lines used in this study were Columbia ecotype. Seedlings were grown at a constant temperature of 22 °C under different light regimes depending on the experiment. The *prmt5-5* mutant was isolated from a pool of transfer DNA (T-DNA) mutagenized plants (stock no. CS31100 from the Arabidopsis Biological Research Center (ABRC)). The mutation causing the circadian phenotype was not linked to a T-DNA and was identified by positional cloning and sequencing. The *flc* (SALK_092716), *atprmt5-1* (SALK_065814) and *atprmt5-2* (SALK_095085) mutants were obtained from ABRC³¹. Seeds were stratified for four days in the dark at 4 °C and then sown onto either soil or solid Murashige and Skoog medium³² containing 1% agarose. ***Drosophila* strains and behavioural assays.** Flies were grown and maintained at 25 °C and kept in a 12:12 h light–dark cycle. Activity was monitored in a light–dark cycle for 4 days. Then, flies were released into constant darkness for 8 days. Data were collected using commercially available *Drosophila* activity monitors (TriKinetics). Rhythmicity and FFT analysis were performed using the ClockLab software (Actimetrics) from data collected in constant darkness. *w¹¹¹⁸* flies used as wild-type controls were provided by the Bloomington Stock Center. The *dart5-1* *Drosophila* line was provided by G. Matera and the *csul^{RM50}* *Drosophila* line was provided by J. Anne. *per⁰¹* and *clk^{irk}* mutants and wild-type *yw* flies were also used. In locomotor activity, flies with a single peak over the significance line in a Chi-squared analysis were scored as rhythmic, which was confirmed by visual inspection of the actograms³³. Data shown were obtained from at least three independent experiments. Statistical analysis included was performed with a Student's *t*-test.

Dissection and immunofluorescence. Flies were entrained in a 12-h light–dark cycle at 25 °C and young (0–3-day old) adult flies were decapitated during the first day in constant darkness at circadian time (CT) 2 and CT 14. Brains were fixed in 4% paraformaldehyde in phosphate buffer and then rinsed three times in PBS–Triton (PT). Brains were then blocked in 7% goat serum in PT for 45 min at 25 °C. After the blocking step tissue was incubated with primary antibodies overnight at 4–8 °C. The primary antibodies used were rabbit anti-PER (1:800, donated by the M. Rosbash laboratory) and rat anti-PDF (1:800, generated in our laboratory). The secondary antibodies used were Cy2-conjugated donkey anti-rabbit and Cy3-conjugated anti-rat (Jacksons Immunoresearch) diluted to a final concentration of 1:250, incubated for 2 h at 25 °C. After staining, brains were washed four times for 15 min each and mounted in 80% glycerol (in PT). Images were taken on a Zeiss Pascal confocal microscope. Nuclear quantification of PER immunoreactivity was performed employing the Image J software.

Circadian rhythm analysis in *Arabidopsis*. For leaf movement analysis, plants were entrained in a light–dark cycle, transferred to continuous white fluorescent light and the position of the first pair of leaves was recorded every hour for 5–6 days³⁴. Period estimates were calculated with Brass 3.0 software (Biological Rhythms Analysis Software System, available from <http://www.amillar.org>) and analysed with the FFT–NLLS suite of programs, as described previously^{35,36}. An ANOVA followed by Tukey's Multiple Comparison Test were used for comparisons among genotypes. For bioluminescence assays, wild-type plants and *prmt5-5* mutants carrying the *CAB2::LUC³⁶* or the *TOC1::LUC* reporter³⁷ were grown in a 12:12 h light–dark cycle of cool white fluorescent light and transferred to continuous white fluorescent light for 5–6 days. Seeds carrying the bioluminescence reporters were provided by S. L. Harmer. Bioluminescence rhythms were detected with a microplate luminometer LB-960 (Berthold Technologies) and analysed with Mikrowin 2000 (version 4.29) software (Mikrotek Laborsysteme). Luminescence from individual traces was analysed with Brass 3.0 software³⁶. The phase distribution of clock-regulated genes was evaluated with Phaser (<http://phaser.cgrb.oregonstate.edu>)³⁸.

Flowering time analysis. Flowering time was estimated by counting the number of rosette leaves at the time of there being a 1-cm-high flower bolt. The experiments were performed in a 16:8 h light–dark cycle at a constant temperature of 22 °C.

qRT–PCR expression analysis. RNA samples were subjected to a DNase treatment with RQ1 RNase-Free DNase (Promega). cDNA derived from this RNA was synthesized using Invitrogen SuperScript III and oligo-dT primer. The synthesized cDNAs were amplified with FastStart Universal SYBR Green Master (Roche) using the 7500 Real Time PCR System (Applied Biosystems) cyclers. *ACTIN 8* and *ribosomal protein 49* (*rp49*) transcripts were used as normalization controls for *Arabidopsis* and *Drosophila*, respectively.

***Arabidopsis* gene expression microarray analysis.** Total RNA was extracted from aerial tissue of plants grown in constant white fluorescent light at 22 °C, without prior entrainment to light–dark or temperature cycles. Each sample consisted of 3–4 plants, to reduce biological variation. Five micrograms of RNA from 21-day-old plants of the wild-type and *prmt5-5* mutant genotypes was processed and hybridized to Affymetrix GeneChip *Arabidopsis* ATH1 Genome Arrays

according to the manufacturer's instructions. Microarray signals were made comparable by scaling the average overall signal intensity of all probe sets to a target signal of 250. Data were analysed using MASS5. We considered for further analysis only those genes having a present call in all samples of either wild-type or *prmt5-5* mutant plants. The statistical significance of signal differences was evaluated by SAM software³⁹. We performed a two-class unpaired *t* test on the signal values, and 100 permutations were used to estimate the false discovery rate. Setting the delta cutoff value at 1.2, and establishing a minimum fold change of 2 between genotypes, 171 genes were found to be significantly regulated by PRMT5 with a false discovery rate of 3.08%.

Whole genome tiling array analysis. For *Arabidopsis* tiling arrays, plants were grown under continuous light for three weeks without prior light or temperature entrainment. For *Drosophila* tiling arrays, flies were grown and maintained at 25 °C in vials containing standard cornmeal medium, and kept in a 12:12 h light–dark cycle and transferred to constant darkness. Head collection and homogenization was done every three hours during the first day in constant darkness. RNA samples from different time points throughout a circadian cycle were pooled together and treated as one. GeneChip *Arabidopsis* Tiling 1.0R Array (Affymetrix) and GeneChip *Drosophila* Tiling 1.0R Array (Affymetrix) were processed according to standard Affymetrix protocols. For *Drosophila* array annotation, perfect match probes from the *Drosophila* tiling 1.0R array (Affymetrix) were megablasted against the *Drosophila* genome release version 5 with word size ≥8 and *E* value ≤0.01. Single perfect matches, without a second partial match of >18/25 bp, were selected, giving a total of 2,930,433 unique probes. These were mapped to annotated mRNAs as exon, intron, intergenic, or flanking probes that span an annotated boundary. Intergenic probes, flanking probes, and probes interrogating multiple transcripts were excluded from the *Drosophila* RNA data. The *Arabidopsis* tiling 1.0R array was annotated previously⁴⁰. For both *Arabidopsis* and *Drosophila*, gene expression and splicing analysis was performed as previously described⁴⁰. Log transformed intensity data from wild types and mutants were fitted with a linear model: intensity = genotype + error, where the genotype term contrasts wild types and mutants. Exonic splicing analysis used probe intensities corrected by mean gene expression. FDRs were determined by 20 permutations⁴⁰.

Bioinformatic analysis of donor splice-site sequences. The genome-wide frequency of each nucleotide for each position of the 5'-splice-site sequence was obtained from SpliceRack⁴¹. The over- or underrepresentation of a particular nucleotide relative to its genome-wide frequency and *P* value was calculated using the hypergeometric test. Nucleotide frequencies were represented using the WebLogo software⁴².

Radioactive PCR for *PRR9* and *per* splicing assessment. Radioactive PCR amplification was performed containing 1.5 U of Taq polymerase (Invitrogen). Primers used for *per* amplification were taken from ref. 5. RT–PCR products were electrophoresed and detected by autoradiography. Radioactivity amounts per band were measured in a scintillation counter (Cerenkov method).

High-resolution RT–PCR panels. Plants were grown under continuous light for three weeks without prior light or temperature entrainment and RNA samples were obtained as described earlier. First-strand cDNA synthesis, PCR protocol and splicing analysis were previously described²¹.

Western blot analysis. Grind tissue in liquid nitrogen and resuspend the powder in 200–500 µl of Lysis Buffer (50 mM Tris–HCl, pH 7.5, 150 mM NaCl, 1% Nonidet P-40, 0.5% deoxycholate and protease inhibitors; Roche) or RIPA buffer. Sample loading was made with 2× Laemmli's buffer in 15–18% SDS–polyacrylamide gel electrophoresis gels. After transference to PVDF membranes the immunoblotting was made with Y12 and SYM10 antibodies according to manufacturer's recommendations.

Sequential chromatin immunoprecipitation (Re–ChIP) assays. Two-week-old seedlings were grown in a 12:12 h light–dark cycle and treated as previously described⁴³. Before the elution from the beads, the immunoprecipitated complexes were released with 10 mM dithiothreitol and then diluted 20 times with Re–ChIP Buffer (20 mM Tris–HCl, pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 1 mM phenylmethylsulphonyl fluoride, 1 µg ml^{−1} aprotinin, 1 µg ml^{−1} pepstatin A, 1 µg ml^{−1} antipain^{−1}, 1 µg ml^{−1} leupeptin and 1 µg ml^{−1} chymostatin) and subjected again to the ChIP procedure (adapted from ref 44). The first immunoprecipitation was performed with anti-histone H4 and the second one (Re–ChIP) with anti-dimethyl-arginine symmetric (SYM10) antibodies (Upstate Biotechnology). DNA obtained was purified using a GFX PCR DNA and Gel Band Purification kit (GE Healthcare) following standard procedures. The amplified promoter region of the *FLC* gene (−523 to −273) was previously described to be subject to epigenetic control⁴⁵. Two promoter regions of the *PRR9* gene were amplified: −621 to −443 and −61 to 83. PCR amplification products were electrophoresed in agarose gel and were stained with SYBR green (Sigma–Aldrich) following the manufacturer's recommendations. Gels were quantitative in the range of DNA concentrations used. Images were captured with the Kodak Digital Science

System, and quantification was performed with ImageQuant software (Molecular Dynamics) and Scion Image software. Here we show one representative of two independent experiments.

31. Alonso, J. M. *et al.* Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**, 653–657 (2003).
32. Murashige, T. & Skoog, F. A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant.* **15**, 473–497 (1962).
33. Rezával, C., Werbach, S. & Ceriani, M. F. Neuronal death in *Drosophila* triggered by GAL4 accumulation. *Eur. J. Neurosci.* **25**, 683–694 (2007).
34. Hicks, K. A. *et al.* Conditional circadian dysfunction of the *Arabidopsis* early-flowering 3 mutant. *Science* **274**, 790–792 (1996).
35. Plautz, J. D. *et al.* Quantitative analysis of *Drosophila period* gene transcription in living animals. *J. Biol. Rhythms* **12**, 204–217 (1997).
36. Millar, A. J., Carre, I. A., Strayer, C. A., Chua, N. H. & Kay, S. A. Circadian clock mutants in *Arabidopsis* identified by luciferase imaging. *Science* **267**, 1161–1163 (1995).
37. Strayer, C. *et al.* Cloning of the *Arabidopsis* clock gene *TOC1*, an autoregulatory response regulator homolog. *Science* **289**, 768–771 (2000).
38. Michael, T. P. *et al.* A morning-specific phytohormone gene expression program underlying rhythmic plant growth. *PLoS Biol.* **6**, e225 (2008).
39. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98**, 5116–5121 (2001).
40. Zhang, X., Byrnes, J., Gal, T., Li, W.-H. & Borevitz, J. Whole genome transcriptome polymorphisms in *Arabidopsis thaliana*. *Genome Biol.* **9**, R165 (2008).
41. Sheth, N. *et al.* Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res.* **34**, 3955–3967 (2006).
42. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
43. Perales, M. & Mas, P. A functional link between rhythmic changes in chromatin structure and the *Arabidopsis* biological clock. *Plant Cell* **19**, 2111–2123 (2007).
44. Métivier, R. *et al.* Estrogen receptor- α directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* **115**, 751–763 (2003).
45. Bastow, R. *et al.* Vernalization requires epigenetic silencing of FLC by histone methylation. *Nature* **427**, 164–167 (2004).

Identification of UBIAD1 as a novel human menaquinone-4 biosynthetic enzyme

Kimie Nakagawa¹, Yoshihisa Hirota¹, Natsumi Sawada¹, Naohito Yuge¹, Masato Watanabe¹, Yuri Uchino¹, Naoko Okuda¹, Yuka Shimomura¹, Yoshitomo Suhara¹ & Toshio Okano¹

Vitamin K occurs in the natural world in several forms, including a plant form, phyloquinone (PK), and a bacterial form, menaquinones (MKs). In many species, including humans, PK is a minor constituent of hepatic vitamin K content, with most hepatic vitamin K content comprising long-chain MKs. Menaquinone-4 (MK-4) is ubiquitously present in extrahepatic tissues, with particularly high concentrations in the brain, kidney and pancreas of humans and rats^{1–3}. It has consistently been shown that PK is endogenously converted to MK-4 (refs 4–8). This occurs either directly within certain tissues or by interconversion to menadione (K₃), followed by prenylation to MK-4 (refs 9–12). No previous study has sought to identify the human enzyme responsible for MK-4 biosynthesis. Previously we provided evidence for the conversion of PK and K₃ into MK-4 in mouse cerebra¹³. However, the molecular mechanisms for these conversion reactions are unclear. Here we identify a human MK-4 biosynthetic enzyme. We screened the human genome database for prenylation enzymes and found *UbiA prenyltransferase containing 1* (UBIAD1), a human homologue of *Escherichia coli* prenyltransferase *menA*. We found that short interfering RNA against the UBIAD1 gene inhibited the conversion of deuterium-labelled vitamin K derivatives into deuterium-labelled-MK-4 (MK-4-d₇) in human cells. We confirmed that the UBIAD1 gene encodes an MK-4 biosynthetic enzyme through its expression and conversion of deuterium-labelled vitamin K derivatives into MK-4-d₇ in insect cells infected with UBIAD1 baculovirus. Converted MK-4-d₇ was chemically identified by ²H-NMR analysis. MK-4 biosynthesis by UBIAD1 was not affected by the vitamin K antagonist warfarin. UBIAD1 was localized in endoplasmic reticulum and ubiquitously expressed in several tissues of mice. Our results show that UBIAD1 is a human MK-4 biosynthetic enzyme; this identification will permit more effective decisions to be made about vitamin K intake and bone health.

Vitamin K is a cofactor for γ -glutamyl carboxylase (GGCX), an enzyme that catalyses the post-translational modifications of several vitamin K-dependent proteins, including coagulation factors, osteocalcin and matrix Gla protein¹⁴. In addition, vitamin K is a transcriptional regulator that modulates the expression of genes in bone through the activation of the steroid and xenobiotic receptor (SXR)^{15–17}. Natural vitamin K is found in two forms: a plant form, phyloquinone (PK) and bacterial forms, menaquinones (MKs)¹⁸. All forms of vitamin K have 2-methyl-1,4-naphthoquinone (menadione; K₃) as a common ring structure. At the 3-position of the ring, PK has a phytyl side chain. In contrast, MKs have several repeating prenyl units, this number being given as a suffix (namely menaquinone-*n* (MK-*n*)).

Early work showed that 2-¹⁴C-methyl-1,4-naphthoquinone was converted to 2-¹⁴C-methyl-3-geranylgeranyl-1,4-naphthoquinone in chicks and rats⁴; in chickens and pigeons, PK doubly labelled with ³H in the 2-methyl group of the naphthoquinone nucleus and ¹⁴C in the side chain was converted to MK-4 singly labelled with ³H (refs 5, 6). The use of this doubly labelled PK showed that the ¹⁴C-labelled side chain of PK was cleaved and replaced with an unlabelled geranylgeranyl side

chain. From these experiments it was confirmed that not only PK, but also MK-4 itself, was endogenously converted to MK-4.

With regard to the conversion site, there are two possibilities: one is in the gut by bacteria and the other is in the intestine and/or local tissues. However, the former is unlikely, as two groups^{11,12} proved independently that conversion took place in both germ-free rats and control animals. The latter is more likely, and tissue MK-4 may originate from PK. If this is so, two pathways for the conversion of PK to MK-4 are possible: first, side-chain removal occurs during intestinal absorption and then the released K₃ is transferred to tissues through the bloodstream and is subsequently prenylated to form MK-4; and second, after the transfer of PK to tissues, side-chain cleavage and geranylgeranylation occur simultaneously within tissues.

It has been reported¹⁹ that urinary K₃ excretion increases after single doses of oral PK and mirrors PK intakes in a 30-day depletion/repletion human metabolic study, implicating K₃ as an intermediate in the conversion of PK to MK-4. We recently confirmed that the conversion of PK to MK-4 took place in a slice culture of mouse cerebra and a primary culture of mouse cerebral hemispheres, although we could not measure the putative K₃ intermediate for methodological reasons¹³. Until now, the conversion site and the putative enzyme(s) responsible for the conversion of PK to MK-4 have not been defined.

In this study we sought the enzyme responsible for the conversion of PK to MK-4 in human osteoblast-like MG-63 cells, primarily because previous studies of the biological actions of vitamin K have been performed exclusively in the context of bone metabolism. First, we confirmed that administration of deuterium-labelled PK (PK-d₇), MK-4 (MK-4-d₁₂) or K₃ (K₃-d₈) led to high levels of MK-4-d₇ in mouse bones (Fig. 1a and Supplementary Fig. 1). The MK-4-d₇ tissue distribution pattern after intake of PK-d₇ or MK-4-d₁₂ was comparable to that found after administration of K₃-d₈. Moreover, we found that PK-d₇, MK-4-d₁₂ and K₃-d₈ were converted into MK-4-d₇ and its epoxide in MG-63 cells (Fig. 1b; for PK-d₇, MK-4-d₇ epoxide was detected at concentrations of 0.12 pmol per mg protein). Vitamin K epoxide is a marker of cofactor activity for GGCX, and thus the converted MK-4 acts as a cofactor for GGCX in MG-63 cells.

In *E. coli*, MKs are involved in several anaerobic electron transport systems²⁰: it is the major transporter of electrons under anaerobic growth conditions^{21,22}. Six genes (*menA*, *menB*, *menC*, *menD*, *menE* and *menF*) are involved in the biosynthetic pathway of MKs, and a key reaction for menaquinone biosynthesis in *E. coli* is the conversion of 1,4-dihydroxy-2-naphthoic acid (a bicyclic naphthalenoid) to the membrane-bound demethylmenaquinone. The key enzyme catalysing this reaction is encoded by *menA* (ref. 22).

Several human homologues of the *E. coli* menaquinone biosynthetic enzymes exist in the human genome database. However, the biological functions of these homologous genes remain unknown. UBIAD1 was the first identified mammalian homologue of *E. coli* *menA*. In *E. coli*, the *menA* gene encodes a prenyltransferase that is involved in the vitamin K biosynthetic pathway²³. However, the function of human UBIAD1 has not been determined. Other prenylation enzymes, UbiA

¹Department of Hygienic Sciences, Kobe Pharmaceutical University, 4-19-1, Motoyamakita-machi, Higashinada-ku, Kobe 658-8558, Japan.

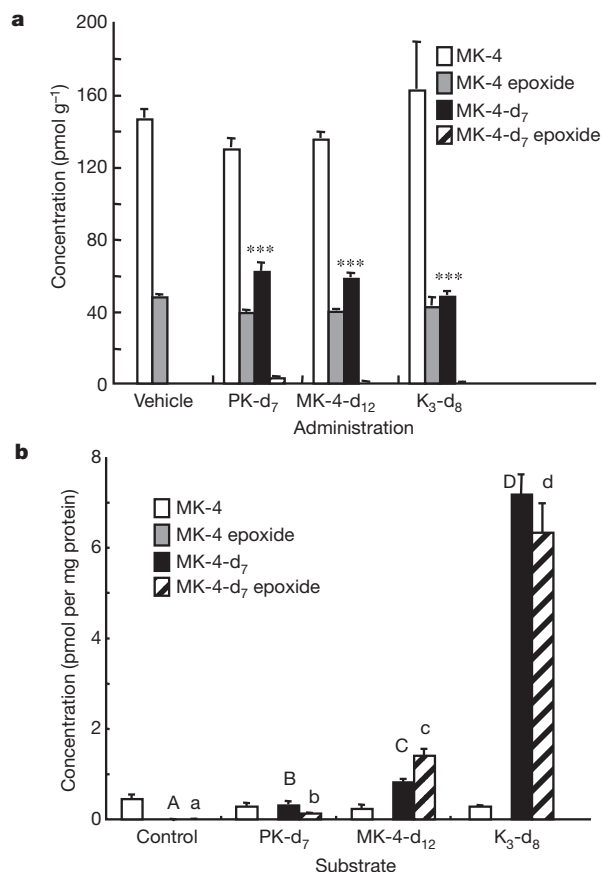


Figure 1 | Conversion activities of deuterium-labelled vitamins K.

a, Accumulation of MK-4-d₇ and MK-4-d₇ epoxides in bones of mice administered with PK-d₇, MK-4-d₁₂ or K₃-d₈. PK-d₇, MK-4-d₁₂ or K₃-d₈ was orally administered at 10 µmol per kg (body weight). After 24 h, concentrations of MK-4-d₇ and its epoxide were measured. Values are means and s.e.m. ($n = 5$). Three asterisks, $P < 0.001$ versus MK-4-d₇ concentration of the vehicle-treated group. **b**, Conversions of PK-d₇, MK-4-d₁₂ or K₃-d₈ into MK-4-d₇ in human osteoblast-like MG-63 cells. MG-63 cells were incubated with vehicle (ethanol), PK-d₇, MK-4-d₁₂ or K₃-d₈ (1 µM) for 24 h, and lipids were extracted for measurement of MK-4-d₇ and its epoxide. Values are means and s.e.m. ($n = 5$). A, B, C and D (or a, b, c and d) in each individual column indicate a significant difference ($P < 0.001$).

in *E. coli* and COQ2 in the yeast *Saccharomyces cerevisiae*, catalyse the prenylation of *p*-hydroxybenzoate, a critical step in ubiquinone biosynthesis.

In humans this step is catalysed by human COQ2, a coenzyme Q2 homologue of prenyltransferase (in yeast)²⁴. The low sequence homology between the *UBIAD1* and *COQ2* genes imply a different role for *UBIAD1*. We therefore predicted that either *UBIAD1* or *COQ2* is a prenylation enzyme for MK-4 biosynthesis in humans. To check this hypothesis, we used short interfering RNA (siRNA) against *UBIAD1* and *COQ2* genes to test their abilities to inhibit the conversions of K₃-d₈ and MK-4-d₁₂ into MK-4-d₇ in MG-63 cells. In contrast with *COQ2* siRNA, *UBIAD1* siRNA strongly decreased the conversions of K₃-d₈ and MK-4-d₁₂ into MK-4-d₇ (Fig. 2a and Supplementary Fig. 2a, b). We also constructed a *UBIAD1* expression vector and analysed the K₃-d₈ into MK-4-d₇ in *UBIAD1* expression-vector-transfected MG-63 cells. MK-4-d₇ levels were correlated with *UBIAD1* messenger RNA levels and *UBIAD1* protein expression levels (Fig. 2b and Supplementary Fig. 3a, b).

To confirm that we had identified the *UBIAD1* gene as relating to a biosynthetic enzyme for MK-4, we expressed the most prevalent form of the enzyme in *Spodoptera frugiperda* (Sf9) cells. *UBIAD1* expression was detected in *UBIAD1* baculovirus-infected and not in control-infected Sf9 cells (Supplementary Fig. 4). Sf9 cells showed no measurable

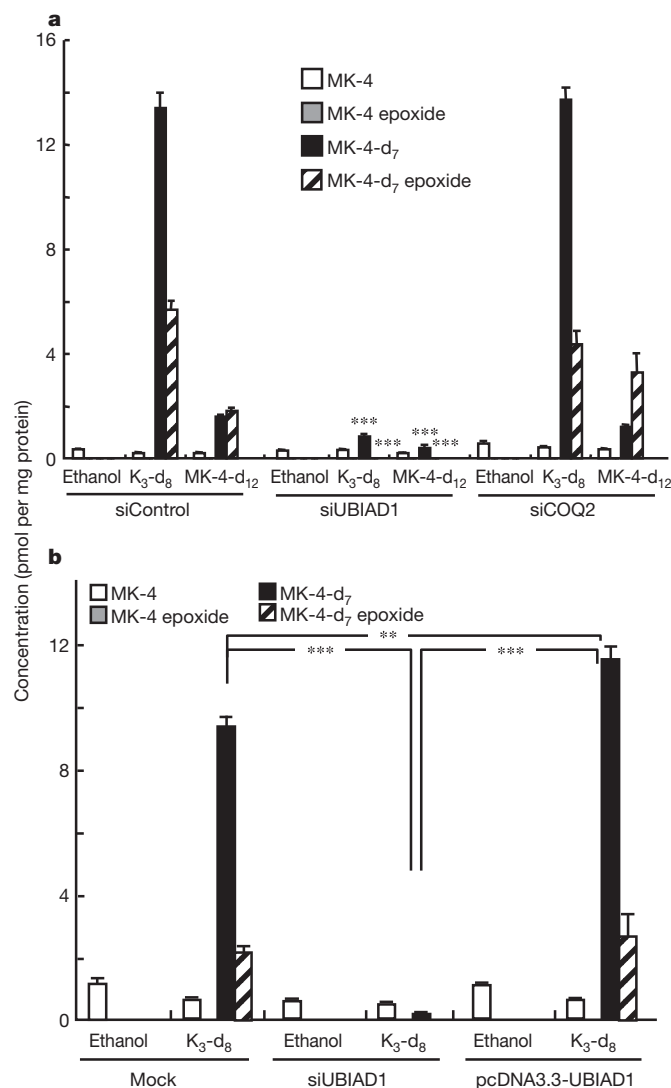


Figure 2 | Conversion activity of UBIAD1 in MG-63 cells. **a**, Conversion of K₃-d₈ or MK-4-d₁₂ into MK-4-d₇ in MG-63 cells transfected with control siRNA (siControl), *UBIAD1* siRNA (siUBIAD1) or *COQ2* siRNA (siCOQ2). Values are means and s.e.m. ($n = 6$). Three asterisks, $P < 0.001$ versus control-siRNA-transfected cells with the same compound treatment. **b**, Conversion of K₃-d₈ into MK-4-d₇ in MG-63 cells transfected with *UBIAD1* siRNA or *UBIAD1* expression vector. Values are means and s.e.m. ($n = 6$). Two asterisks, $P < 0.01$; three asterisks, $P < 0.001$ versus each group with K₃-d₈ treatment.

conversion activity, but showed conversion activity when infected with *UBIAD1* baculovirus (Fig. 3a). *UBIAD1* was able to convert K₃-d₈, PK-d₇ and MK-4-d₁₂ into MK-4-d₇. The comparative conversion ratio of these vitamins K was K₃-d₈ > MK-4-d₁₂ > PK-d₇. These results indicate that *UBIAD1* may cleave the side chain to release K₃ and then prenylate it with geranylgeranyl pyrophosphate (GGPP)²⁵ to form MK-4.

Finally, we obtained 4 µg of the MK-4-d₇ fraction from K₃-d₈-treated *UBIAD1*-expressing Sf-9 cells. From liquid-chromatography/atmospheric-pressure chemical ionization tandem mass spectrometry (LC-APCI-MS/MS) analyses, authentic MK-4-d₇ showed a parent peak (Q1) at $m/z = 452.4$ [$M + H$]⁺ and a product ion peak (Q3) at $m/z = 194.3$. The MK-4-d₇ fraction showed the same parent peak at $m/z = 452.4$ and a product ion peak at $m/z = 194.3$ (Fig. 3b). Figure 3c shows the ²H-NMR spectra of authentic MK-4-d₇ and MK-4-d₇ fractions. The resonance values derived from the deuterium-labelled 2-methyl-1,4-naphthoquinone ring of the MK-4-d₇ fraction coincided with those of authentic MK-4-d₇. Together, these results show that *UBIAD1* is a human biosynthetic enzyme for MK-4.

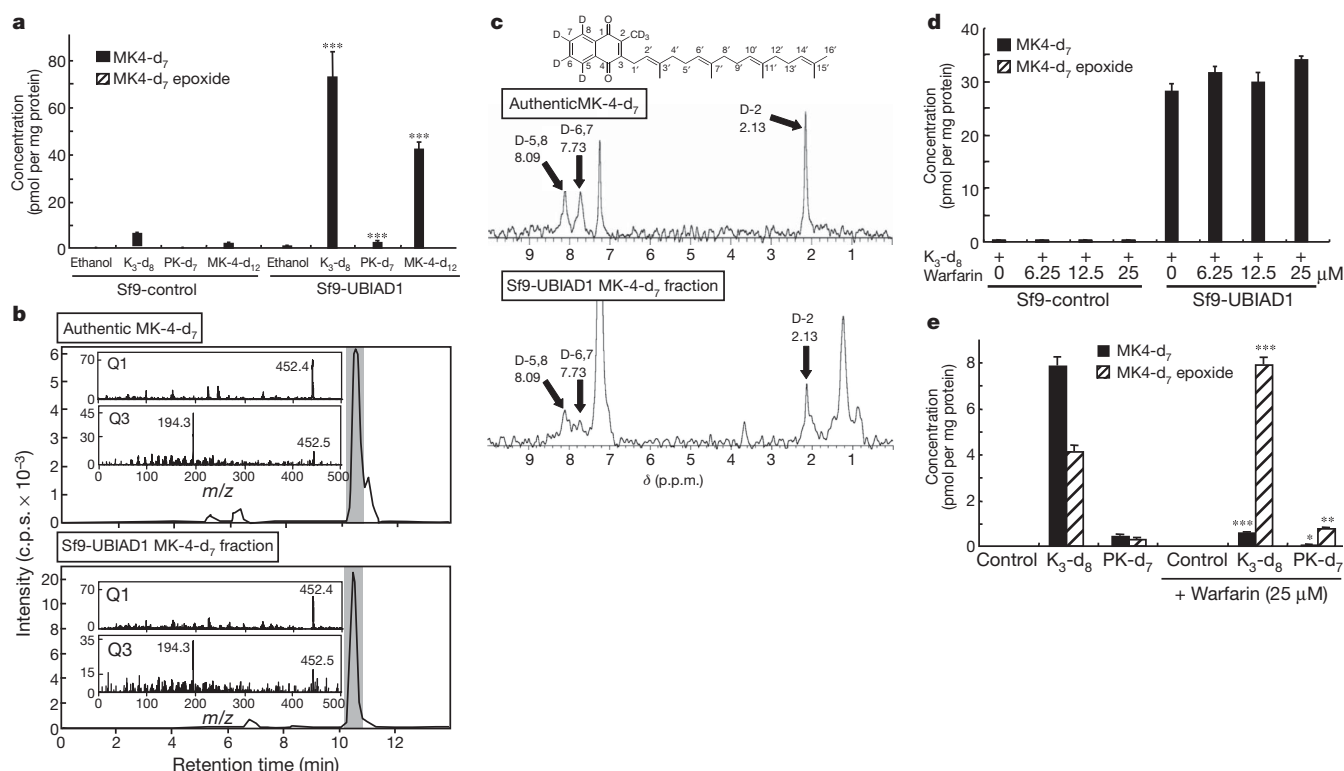


Figure 3 | MK-4 biosynthetic activity of UBIAD1 in Sf9 cells and effects of warfarin on MK-4 biosynthesis. **a**, Conversions of K_3 - d_8 , PK- d_7 or MK-4- d_{12} into MK-4- d_7 in UBIAD1 baculovirus-infected Sf9 cells. Values are means and s.e.m. ($n = 6$). Three asterisks, $P < 0.001$ versus control-infected Sf9 cells with the same compound treatment. **b**, LC-APCI-MS/MS analyses of the MK-4- d_7 fraction purified from UBIAD1 baculovirus-infected Sf9 cells treated with K_3 - d_8 . The shaded portions of the graphs indicate the MK-4- d_7 fraction.

It was previously reported that the accumulation of MK-4 in the organs of rats fed with 6,7- 3H labelled K_3 was markedly decreased by the vitamin K antagonist dicoumarol, which has a similar mechanism of action to that of warfarin, such that in the kidney and heart MK-4 concentrations were decreased 6-fold and 20-fold, respectively²⁶. It remains unclear whether this antagonistic effect is due to an inhibition of the conversion reaction or whether it is due to other mechanisms, such as MK-4 being used by GGCX and accumulated as MK-4 epoxide^{11,27}. To determine the effect of vitamin K antagonists such as warfarin on the conversion activity of UBIAD1, we performed additional experiments with UBIAD1-expressing Sf9 and MG-63 cells. The conversion of K_3 - d_8 to MK-4- d_7 was not affected by the treatment of UBIAD1-expressing Sf9 cells with warfarin (Fig. 3d). In MG-63 cells, warfarin decreased the concentration of MK-4- d_7 , although MK-4- d_7 epoxide accumulated (Fig. 3e). These results indicate that warfarin does not affect MK-4 biosynthesis by UBIAD1, but only induces the accumulation of MK-4 epoxide.

Next we examined the localization of the UBIAD1 enzyme in MG-63 cells. The UBIAD1 gene is 22 kilobases in length, with up to five exons. On the basis of its amino-acid structure, the UBIAD1 enzyme is speculated to be localized in the endoplasmic reticulum (ER). To study the subcellular localization of UBIAD1, we generated constructs expressing green fluorescent protein (GFP)-tagged UBIAD1 fusion proteins (UBIAD1-GFP) for the stable transfection of MG-63 cells. Figure 4a shows that the green fluorescence of UBIAD1-GFP decorated the mesh-like structure of the ER in the cytoplasm and co-localized exactly with the red ER marker label (ER-tracker Red) but did not co-localize with the Golgi marker (4,4-difluoro-4-bora-3a,4a-diaza-s-indacene (BODIPY)-TR ceramide). These results strongly indicate that UBIAD1 is a biosynthetic enzyme for MK-4 located in the ER.

We also examined the distributions of *Ubiad1* mRNA in mouse tissues and found that mouse *Ubiad1* mRNA was expressed ubiquitously

in all tissues tested (Supplementary Fig. 5). We measured the concentrations of MK-4, MK-4- d_7 and their epoxides in several tissues, including the cerebrum, liver and pancreas of mice orally administered with PK- d_7 . MK-4- d_7 biosynthetic activity was correlated with mouse *Ubiad1* mRNA expression in tissues, except for the heart (Supplementary Fig. 6a, b). Further, the tissue accumulation patterns of MK-4- d_7 and its epoxide coincided with those of endogenous MK-4 and its epoxide (Supplementary Fig. 6b, c). The expression of *Ubiad1* mRNA in heart was highest in the tissues tested, whereas the concentrations of MK-4, MK-4- d_7 and their epoxides were relatively low. This may suggest an unknown function of UBIAD1 in addition to MK-4 biosynthesis. Taken together, our results strongly suggest that PK is converted to MK-4 by UBIAD1 and accumulates as MK-4 in tissues.

To ascertain the mechanism for the side-chain cleavage and prenylation activities of UBIAD1, we measured UBIAD1 activity by using the microsomal fraction of UBIAD1 baculovirus-infected Sf9 cells (Supplementary Fig. 7). We detected the activity of UBIAD1 in converting PK- d_7 and K_3 - d_8 to MK-4- d_7 in the microsomes (Fig. 4b and Supplementary Fig. 8). Moreover, the prenylation activity of UBIAD1 was proportional to the concentration of the side-chain substrate (GGPP). These results suggest that UBIAD1 is a novel biosynthetic enzyme for MK-4 that may have both side-chain cleavage and prenylation activities (Fig. 4d). Because UBIAD1 is a ubiquitously expressed enzyme and shows its MK-4 biosynthesis activity in most tissues, both MK-4 and UBIAD1 may be physiologically essential factors. Several studies have recently reported that UBIAD1 is a candidate gene for Schnyder crystalline corneal dystrophy (SCCD), an autosomal dominant disorder involving deposition of cholesterol in the corneal stroma^{28–30}. Although there have been studies on UBIAD1 mutations in SCCD, the function of UBIAD1 was unknown. Our findings imply that the MK-4 biosynthesis activity of UBIAD1 may be related to this

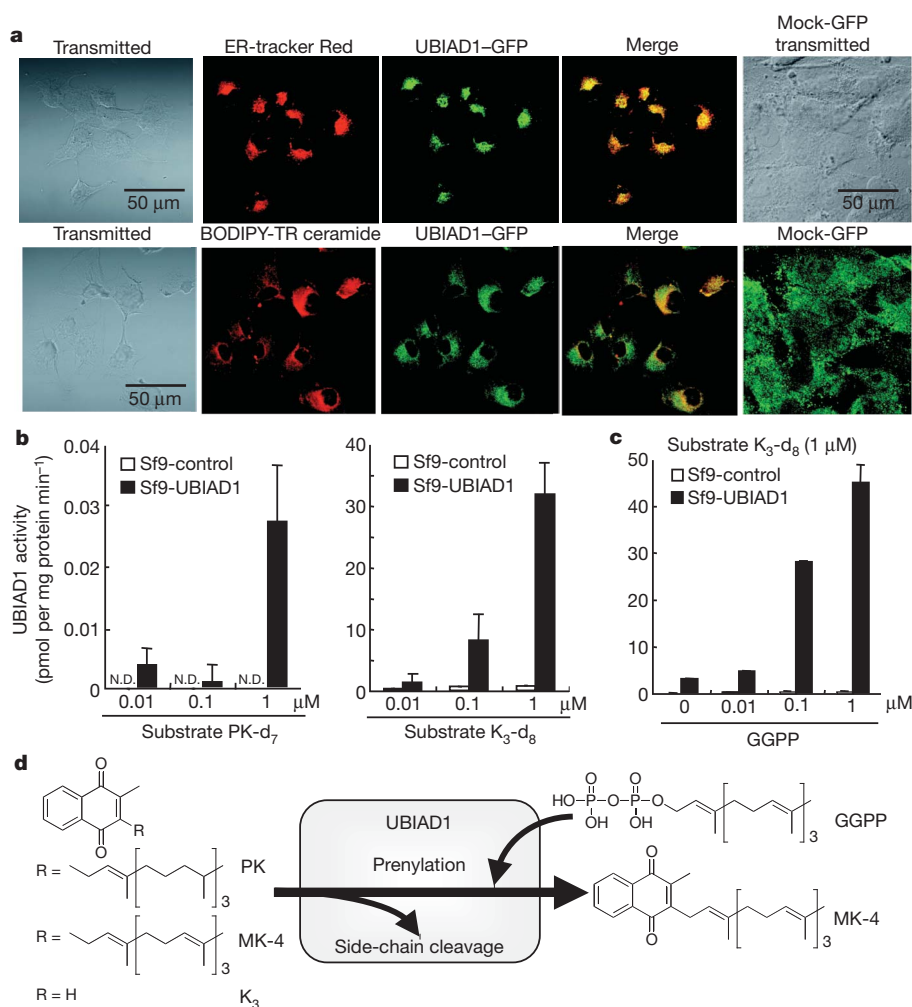


Figure 4 | Subcellular localization of UBIAD1 in MG-63 cells, UBIAD1 activity in microsomes from *UBIAD1* baculovirus-infected Sf9 cells, and scheme of UBIAD1 activity. **a**, MG-63 cells stably transfected with a UBIAD1-GFP expression vector were stained with ER-tracker Red or BODIPY-TR ceramide (red) and were detected by GFP fluorescence (green). Merged images of GFP fluorescence and by ER-marker or Golgi-marker fluorescence are shown at the right. The control construct (mock-GFP) showed

a diffuse fluorescence throughout the cytoplasm. **b**, Activity of UBIAD1 in microsomes prepared from *UBIAD1* baculovirus-infected Sf9 cells. Values are means and s.e.m. ($n = 6$). **c**, GGPP dose-dependent prenylation activity of UBIAD1 in microsomes prepared from *UBIAD1* baculovirus-infected Sf9 cells. Values are means and s.e.m. ($n = 6$). **d**, Scheme showing the conversion of vitamin K derivatives to MK-4 by UBIAD1.

disorder. To address these issues, enzymological characteristics of UBIAD1 and the molecular mechanisms regulating *UBIAD1* gene expression need to be elucidated.

METHODS SUMMARY

Deuterium-labelled vitamin K compounds were synthesized^{13,25}. To assess the conversions of PK-d₇, MK-4-d₁₂ and K₃-d₈ into MK-4-d₇, MK-4-d₇ in bone extracted from mice after 24 h of administration of deuterium-labelled vitamins K was measured by LC-APCI-MS/MS. Human osteoblast-like MG-63 cells were provided by K. Ozono. MG-63 cells were maintained in DMEM medium (10% fetal calf serum (FCS)) and transfected with siRNA against *UBIAD1* or *COQ2* and *UBIAD1* expression vectors. These cells were treated with K₃-d₈, PK-d₇ or MK-4-d₁₂ (1 μM) for 24 h. MK-4-d₇ and MK-4-d₇ epoxides in cells were measured by LC-APCI-MS/MS. The complementary DNA for the UBIAD1 coding region was cloned into pLex/Bac-1 (Merck) with Insect GeneJuice (Merck) in Sf9 insect cells. *UBIAD1* baculovirus-infected Sf9 cells were treated for 24 h with deuterium-labelled vitamins K. These vitamin K concentrations were measured by LC-APCI-MS/MS. MK-4-d₇ converted from K₃-d₈ in infected Sf9 cells was identified by ²H-NMR analysis. To investigate localization of UBIAD1 in cells, GFP-tagged UBIAD1 vectors were constructed and stably transfected in MG-63 cells. Cloned MG-63-UBIAD1-GFP cells were stained with ER-tracker Red or BODIPY-TR ceramide (Molecular Probes). Fixed cells were viewed under an Olympus IX70 microscope and photographed. *UBIAD1* mRNA expression in mouse tissues was measured by quantitative real-time RT-PCR. Microsomes of

UBIAD1 baculovirus-infected Sf9 cells were incubated for 3 h with deuterium-labelled vitamins K in reaction buffer containing 0.01–1 μM GGPP, 10 mM MgCl₂ and 1 mM dithiothreitol in 100 mM Tris-HCl pH 7.8. The concentrations of vitamins K were measured by LC-APCI-MS/MS.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 April; accepted 27 August 2010.

Published online 17 October 2010.

1. Shearer, M. J., McCarthy, P. T., Crampton, O. E. & Mattock, M. B. in *Current Advances in Vitamin K Research* (ed. Suttie, J. W.) 437–452 (Elsevier, 1988).
2. Thijssen, H. H. W. & Drittij-Reijnders, M. J. Vitamin K distribution in rat tissues: dietary phylloquinone is a source of tissue menaquinone-4. *Br. J. Nutr.* **72**, 415–425 (1994).
3. Thijssen, H. H. W. & Drittij-Reijnders, M. J. Vitamin K status in human tissues: tissue-specific accumulation of phylloquinone and menaquinone-4. *Br. J. Nutr.* **75**, 121–127 (1996).
4. Martius, C. & Esser, H. O. Über die Konstitution des im Tierkörper aus Methylnaphthochinon gebildeten K-Vitamins. *Biochem. Z.* **331**, 1–9 (1958).
5. Billeter, M. & Martius, C. Über die umwandlung von phyllochinon (vitamin K₁) in vitamin K₂ (20) in tierkörper. *Biochem. Z.* **333**, 430–439 (1960).
6. Martius, C. The metabolic relationships between the different K vitamins and the synthesis of the ubiquinones. *Am. J. Clin. Nutr.* **9**, 97–103 (1961).
7. Guillaumont, M. *et al.* Hepatic concentration of vitamin K active compounds after application of phylloquinone to chickens on a vitamin K deficient or adequate diet. *Int. J. Vitam. Nutr. Res.* **62**, 15–20 (1992).

8. Thijssen, H. H. W., Drittij-Reijnders, M. J. & Fischer, M. A. J. G. Phylloquinone and menaquinone-4 distribution in rats: synthesis rather than uptake determines menaquinone-4 organ concentrations. *J. Nutr.* **126**, 537–543 (1996).
9. Shearer, M. J. & Newman, P. Metabolism and cell biology of vitamin K. *Thromb. Haemost.* **100**, 530–547 (2008).
10. Dialameh, G. H., Yekundi, K. G. & Olson, R. E. Enzymatic alkylation of menaquinone-0 to menaquinones by microsomes from chick liver. *Biochim. Biophys. Acta* **223**, 332–338 (1970).
11. Ronden, J. E., Drittij-Reijnders, M. J., Vermeer, C. & Thijssen, H. H. Intestinal flora is not an intermediate in the phylloquinone–menaquinone-4 conversion in the rat. *Biochim. Biophys. Acta* **1379**, 69–75 (1998).
12. Davidson, R. T., Foley, A. L., Engelke, J. A. & Suttie, J. W. Conversion of dietary phylloquinone to tissue menaquinone-4 in rats is not dependent on gut bacteria. *J. Nutr.* **128**, 220–223 (1998).
13. Okano, T. *et al.* Conversion of phylloquinone (vitamin K₁) into menaquinone-4 (vitamin K₂) in mice: two possible routes for menaquinone-4 accumulation in cerebra of mice. *J. Biol. Chem.* **283**, 11270–11279 (2008).
14. Suttie, J. W. in *Present Knowledge in Nutrition* 7th edn (eds Ziegler, E. E. & Filer, L. J. Jr) 137–145 (ILSI Press, 1996).
15. Tabb, M. M. *et al.* Vitamin K₂ regulation of bone homeostasis is mediated by the steroid and xenobiotic receptor SXR. *J. Biol. Chem.* **278**, 43919–43927 (2003).
16. Ichikawa, T., Horie-Inoue, K., Ikeda, K., Blumberg, B. & Inoue, S. Steroid and xenobiotic receptor SXR mediates vitamin K₂-activated transcription of extracellular matrix-related genes and collagen accumulation in osteoblastic cells. *J. Biol. Chem.* **281**, 16927–16934 (2006).
17. Igarashi, M. *et al.* Vitamin K induces osteoblast differentiation through pregnane X receptor-mediated transcriptional control of the *Msx2* gene. *Mol. Cell. Biol.* **27**, 7947–7954 (2007).
18. Booth, S. L. & Suttie, J. W. Dietary intake and adequacy of vitamin K. *J. Nutr.* **128**, 785–788 (1998).
19. Thijssen, H. H., Vervoort, L. M., Schurgers, L. J. & Shearer, M. J. Menadiene is a metabolite of oral vitamin K. *Br. J. Nutr.* **95**, 260–266 (2006).
20. Wallace, B. J. & Young, I. G. Role of quinones in electron transport to oxygen and nitrate in *Escherichia coli*. Studies with a *ubiA⁻ menA⁻* double quinone mutant. *Biochim. Biophys. Acta* **461**, 84–100 (1977).
21. Lin, E. C. C. & Kuritzkes, D. in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology* (eds Neidhardt, F. C. *et al.*) 202–221 (American Society for Microbiology, 1987).
22. Meganathan, R. in *Escherichia coli and Salmonella: Cellular and Molecular Biology* 2nd edn (eds Neidhardt, F. C. *et al.*) 642–656 (American Society for Microbiology, 1996).
23. Suvarna, K., Stevenson, D., Meganathan, R. & Hudspeth, M. E. S. Menaquinone (vitamin K₂) biosynthesis: localization and characterization of the *menA* gene from *Escherichia coli*. *J. Bacteriol.* **180**, 2782–2787 (1998).
24. López-Martín, J. M. *et al.* Missense mutation of the *COQ2* gene causes defects of bioenergetics and *de novo* pyrimidine synthesis. *Hum. Mol. Genet.* **16**, 1091–1097 (2007).
25. Suhara, Y., Wada, A. & Okano, T. Elucidation of the mechanism producing menaquinone-4 in osteoblastic cells. *Bioorg. Med. Chem. Lett.* **19**, 1054–1057 (2009).
26. Taggart, W. V. & Matschiner, J. T. Metabolism of menadiene-6,7-³H in the rat. *Biochemistry* **8**, 1141–1146 (1969).
27. Spronk, H. M. *et al.* Tissue-specific utilization of menaquinone-4 results in the prevention of arterial calcification in warfarin-treated rats. *J. Vasc. Res.* **40**, 531–537 (2003).
28. Orr, A. *et al.* Mutations in the *UBIAD1* gene, encoding a potential prenyltransferase, are causal for Schnyder crystalline corneal dystrophy. *PLoS ONE* **2**, e685 (2007).
29. Weiss, J. S. *et al.* Mutations in the *UBIAD1* gene on chromosome short arm 1, region 36, cause Schnyder crystalline corneal dystrophy. *Invest. Ophthalmol. Vis. Sci.* **48**, 5007–5012 (2007).
30. Weiss, J. S. *et al.* Genetic analysis of 14 families with Schnyder crystalline corneal dystrophy reveals clues to *UBIAD1* protein function. *Am. J. Med. Genet. A* **146**, 271–283 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Sugiura and A. Takeuchi for technical support with ²H-NMR analysis and LC-APCI-MS/MS analysis. This work was supported in part by priority areas from the Ministry of Education, Culture, Sports, Science and Technology (to K.N. and T.O.).

Author Contributions K.N. and T.O. planned the project and analysed the experiments, together with Y.H., N.S., N.Y., M.W., Y.U., N.O., Y.S. and Y.S. The manuscript was written by K.N. and T.O., and all authors commented on it.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.O. (t-okano@kobepharmaceutical-u.ac.jp).

METHODS

Materials. Deuterium-labelled menadione (K_3 -d₈) was purchased from C/D/N Isotopes, Inc. PK epoxide, MK-4 epoxide, ¹⁸O-labelled PK and MK-4 (PK-¹⁸O and MK-4-¹⁸O), deuterium-labelled PK and PK epoxide (PK-d₇ and PK-d₇ epoxide), deuterium-labelled MK-4 and MK-4 epoxide (MK-4-d₇, MK-4-d₁₂ and MK-4-d₇ epoxide) were synthesized in our laboratory, as reported previously^{13,25}.

Comparison of the accumulation of MK-4-d₇ in bones of mice administered with K₃-d₈, PK-d₇ and MK-4-d₁₂. Five 8-week-old female C57BL/6 mice (SLC, Inc.) in each group were given K₃-d₈, PK-d₇ or MK-4-d₁₂ orally as a single dose at 10 μmol per kg (body weight). After 24 h, mice were killed; bones were removed and stored at -80 °C for analysis. Measurements for MK-4-d₇ and MK-4-d₇ epoxide in bone (femur and tibia) used the LC-APCI-MS/MS method described previously¹³.

Comparison of the conversion of K₃-d₈, PK-d₇ or MK-4-d₁₂ into MK-4-d₇ in human osteoblast-like MG-63 cells. Human osteosarcoma MG-63 cells were maintained in DMEM medium (Nakalai Tesque) supplemented with 1% penicillin, 1% streptomycin and 10% FCS (Gibco BRL). MG-63 cells were cultured for 2 days on six-well tissue culture plates (10⁶ cells per well) and treated for 24 h with the culture medium containing K₃-d₈, PK-d₇ or MK-4-d₁₂ (1 μM). Cells were collected and washed three times with cold PBS (Ca²⁺ and Mg²⁺ free) and then stored at -30 °C. After being warmed to 20–25 °C, cells were lysed in 1 ml of PBS(-). Cell lysates (20 μl) were analysed for protein concentrations. ¹⁸O-labelled PK and ¹⁸O-labelled MK-4 were added as internal standards to the cell lysates in brown screw-capped tubes. Measurements for MK-4-d₇ and MK-4-d₇ epoxide in cells used the LC-APCI-MS/MS method described above.

Stealth siRNA treatment. Stealth siRNA *siUBIAD1* is a 25-base-pair (bp) duplex oligoribonucleotide with a sense strand corresponding to nucleotides 619–643 of the reported human *UBIAD1* mRNA sequence. Stealth siRNA *siCOQ2* is a 25-bp duplex oligoribonucleotide with a sense strand corresponding to nucleotides 1016–1040 of the reported human *COQ2* mRNA sequence. Alexa-labelled siRNA (Invitrogen) was used as control siRNA. MG-63 cells were transfected with 50 pmol of each siRNA with the use of Lipofectamine RNAiMAX (Invitrogen) in 1 ml of Opti-MEM serum-reduced medium (Gibco BRL) in accordance with the manufacturer's instructions.

Transfection of human UBIAD1 expression plasmid. A human *UBIAD1* cDNA fragment from the pENTR221-*UBIAD1* plasmid (Invitrogen) was inserted into the TA cloning site of the pcDNA3.3-TOPO-TA cloning vector (Invitrogen) to create pcDNA3.3-*UBIAD1*. MG-63 cells (2 × 10⁵ cells per well) were cultured for 24 h and transfected with 1 μg of pcDNA3.3-*UBIAD1* using the Tfx-50 reagent (Promega).

Real-time quantitative PCR. Total RNA of MG-63 cells or mouse tissues was isolated with Isogen (Nippon Gene) in accordance with the manufacturer's protocol. First-strand cDNA was made with AMV reverse transcriptase (TaKaRa). cDNAs were mixed with a SYBR Green core reagent kit (PE Biosystems), and real-time PCR used the CFX96 real-time PCR system (Bio-Rad). We used human *UBIAD1* (GenBank accession number NM_013319; forward primer 571–590 base pairs; reverse primer 868–887), human *COQ2* (GenBank accession number NM_015697; forward primer 889–908; reverse primer 1004–1023), human β-actin-specific primer sets (GenBank accession number BC013835; forward primer 879–901; reverse primer 1069–1092), mouse *Ubiad1* (GenBank accession number NM_027873; forward primer 1575–1595; reverse primer 1754–1774) and mouse β-actin (GenBank accession number X03672; forward primer 250–271; reverse primer 305–326).

Western blotting. *UBIAD1* expression levels were detected by western blot analysis. The *UBIAD1* antibody was a *UBIAD1*-specific affinity-purified polyclonal antibody raised in rabbits against a *UBIAD1*-specific peptide (CPEQDRLPQRSWRQK-COOH) (MRL Co., Ltd). The peroxidase-conjugated secondary antibody was rabbit immunoglobulin raised in donkey (Santa Cruz) for 1.5 h and *UBIAD1* protein was detected with an electrochemiluminescent detection system (Nakalai Tesque).

Expression of UBIAD1 in Sf9 cell culture and conversion assay. The cDNA for the *UBIAD1* coding region was cloned into pEx/Bac-1 (Merck) with Insect GeneJuice (Merck) in Sf9 insect cells. Recombinant baculovirus was generated with the BacMagic system (Merck) in accordance with the manufacturer's protocol. Sf9 cells were grown at 28 °C in a tissue culture flask in Grace's insect medium (Invitrogen) supplemented with 10% FCS. To achieve the expression of *UBIAD1*, cell cultures (2 × 10⁶ cells ml⁻¹) were infected at a multiplicity of infection of 10 followed by a 72-h expression period. To create a negative control, cells were infected with the pEx/Bac-1 control virus. Infected Sf9 cells were plated for three days on six-well tissue culture plates at a density of 10⁶ cells per well and treated for 24 h with a culture medium containing K₃-d₈, PK-d₇ or MK-4-d₁₂ (1 μM). Measurements for MK-4-d₇ in the cells used the LC-APCI-MS/MS method described above. In LC-APCI-MS/MS analysis, multiple reaction monitoring (MRM) chromatograms and mass spectra of authentic MK-4 and the MK-4 fraction (*m/z* = 452.4 for MK-4-d₇), respectively. The parent and product ion mass spectra (Q1 and Q3, respectively) derived from Q1 were recorded at the time corresponding to the top of the peak (shaded area) on the MRM chromatograms (Fig. 3b).

Identification of MK-4-d₇ in UBIAD1-expressing Sf9 cells. A sample of MK-4-d₇ in K₃-d₈-treated *UBIAD1*-infected Sf9 cells was extracted and purified with procedures described previously⁴. The 500-MHz ¹H-NMR and ²H-NMR spectra of the putative MK-4-d₇ were measured on a Varian VNS-500 (H, 500 MHz; ²H, 77 MHz). The sample was dissolved in 40 μl of CHCl₃ as a nanoprobe for ²H-NMR spectrometry. The number and letter (D) in each spectrum refer to the chemical shift (p.p.m.) and the respective position of the deuterium in the 2-methyl-1,4-naphthoquinone ring or the side chain of MK-4-d₇.

Effects of warfarin on MK-4-d₇ biosynthesis. *UBIAD1*-infected Sf9 cells were plated for two days on six-well tissue culture plates at a density of 10⁶ cells per well and pretreated for 24 h with a culture medium containing warfarin (25 μM). On the next day, cells were treated for 24 h with a culture medium containing K₃-d₈, PK-d₇ or MK-4-d₁₂ (1 μM) with warfarin. MG-63 cells were plated for three days on six-well tissue culture plates at a density of 10⁶ cells per well and pretreated for 24 h with a culture medium containing warfarin (25 μM). On the next day, MG-63 cells were treated for 24 h with a culture medium containing K₃-d₈ or PK-d₇ (1 μM) with warfarin. Measurements for MK-4-d₇ and MK-4-d₇ epoxide in the cells used the LC-APCI-MS/MS method described above.

Localization of GFP-tagged human UBIAD1. A human *UBIAD1* cDNA fragment was inserted into the recombination region of the pcDNA-DEST53 cloning vector (GFP-tagged expression vector; Invitrogen) to create pcDNA-*UBIAD1*-GFP. MG-63 cells were transfected with 10 μg of pcDNA-*UBIAD1*-GFP using the Tfx-50 reagent and cloned in a selective medium containing 500 μg ml⁻¹ G418 (Roche). Cloned MG-63-*UBIAD1*-GFP cells were harvested onto a slide glass with flexiPERM 4-CHAMBER DISC (E&K Scientific), cultured for two days and stained with ER-tracker Red or BODIPY-TR ceramide (Molecular Probe). Cells were fixed with 10% formaldehyde solution, viewed under an Olympus IX70 microscope and photographed.

In vitro UBIAD1 activity. *UBIAD1* activity was characterized with the use of microsomes from *UBIAD1* baculovirus--infected Sf9 cells. The washed cell pellets were resuspended in homogenization buffer containing 100 mM Tris-HCl pH 7.8, 10 mM MgCl₂ and protease inhibitor cocktail (Complete mini, EDTA free; Roche), disrupted by freeze-thawing and sonication on ice and centrifuged at 3,000g for 10 min to remove unbroken cells and nuclei. The cell lysate was centrifuged at 1,00,000g for 60 min. The resulting pellet was resuspended in homogenization buffer and used as microsomes. Reaction mixtures contained deuterium-labelled vitamins K (PK-d₇ or K₃-d₈), 0.01–1 μM GGPP, 10 mM MgCl₂ and 1 mM dithiothreitol in 100 mM Tris-HCl pH 7.8 and 200 mg of microsome protein. Reactions were stopped after 3 h by the addition of ethanol. The resulting mixture was extracted with the same method as for LC-APCI-MS/MS analysis.

CORRIGENDUM

doi:10.1038/nature09548

Ecoenzymatic stoichiometry of microbial organic nutrient acquisition in soil and sediment

Robert L. Sinsabaugh, Brian H. Hill & Jennifer J. Follstad Shah

Nature **462**, 795–798 (2009)

In this Letter, the L/B terms in equations (3) and (4) should be inverted to B/L . The authors report that these equations are not integral to their hypotheses and were not part of the analysis. The sole function of these equations is to provide a heuristic to illustrate that the threshold element ratio (TER) is a distinct parameter from the actual elemental composition of available organic matter.

RETRACTION

doi:10.1038/nature09562

The large-conductance Ca^{2+} -activated K^{+} channel is essential for innate immunity

Jatinder Ahluwalia, Andrew Tinker, Lucie H. Clapp, Michael R. Duchon, Andrey Y. Abramov, Simon Pope, Muriel Nobles & Anthony W. Segal

Nature **427**, 853–858 (2004)

The authors wish to retract this Letter after the report of an inability to reproduce their results¹, later confirmed by another². The studies the authors then conducted led to an internal investigation by University College London, please see the accompanying Supplementary Information for details. The retraction has not been signed by Jatinder Ahluwalia.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

1. Femling, J. K. *et al.* The antibacterial activity of human neutrophils and eosinophils requires proton channels but not BK channels. *J. Gen. Physiol.* **127**, 659–672 (2006).
2. Essin, K. *et al.* Large-conductance calcium-activated potassium channel activity is absent in human and mouse neutrophils and is not required for innate immunity. *Am. J. Physiol. Cell Physiol.* **293**, C45–C54 (2007).

ERRATUM

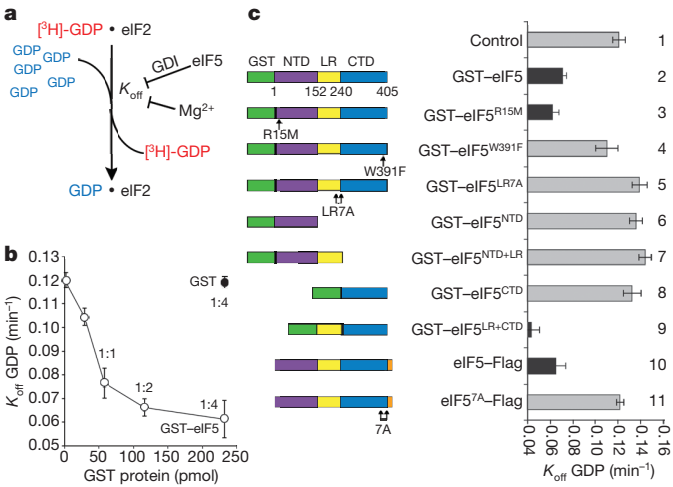
doi:10.1038/nature09550

eIF5 has GDI activity necessary for translational control by eIF2 phosphorylation

Martin D. Jennings & Graham D. Pavitt

Nature 465, 378–381 (2010)

In this Letter, the bar representing the data for Fig. 1c, row 9, was inadvertently deleted. The corrected figure is shown below.



CAREERS

EUROPE Transferable pensions for researchers may soon be a reality **p.125**

Q&A ETH Zurich's president explains the appeal of Switzerland **p.126**

NATUREJOBS For the latest career listings and advice www.naturejobs.com



ACADEMIA

The changing face of tenure

Although still highly desirable, tenure is not as prevalent as it was in some places — and that may not be a bad thing.

BY KAREN KAPLAN

Biologist Rafael Carazo Salas doesn't have tenure — nor is he expecting to pursue the tenure-track system any time soon. As a faculty member at a UK institution, he doesn't have that option — academic tenure per se in the United Kingdom was abolished more than 20 years ago.

But Carazo Salas, a group leader at the University of Cambridge, UK, isn't lying awake at night trying to dream up ways to manoeuvre himself into a tenured or tenure-track research position. Funded by a portable five-year grant from the European Research Council, he is pleased with what he calls a high level of scientific independence conferred by the grant, even though he's well aware that he has no guarantee

of a continuing position at Cambridge at the end of the next four years.

"Everyone would like to have job security," says Carazo Salas, who moved this year from ETH Zurich in Switzerland after his partner secured a Cambridge post. But Carazo Salas is fine with his current position. He may not have job security in perpetuity, but he has autonomy, few administrative duties, and no teaching obligations. "If I secure funds to continue paying my own salary, I can conceivably stay here as long as I want," he says.

Although most academics strive for tenure, experiences such as Carazo Salas's suggest that it is not the only satisfying career course. Early-career academic researchers in the United States, the European Union (EU) and elsewhere are wrestling with major shifts in tenure's definition, availability and value. Seen for decades as the only route to long-term job security and academic freedom, its long-standing symbol as the ultimate prize for academic researchers has been eroding on many fronts. Tenured and tenure-track positions, already hard to secure, have become rarer in some areas because of budget concerns. Other regions are seeing increased interest, as governments and institutions try to attract top talent. Tenure is no longer what it once was, and young scientists might want to survey the features of a changing landscape.

TENURE'S DECLINE

At most North American institutions, tenure is typical for senior faculty appointments such as professors and associate professors. Achieving tenure generally requires a strong record of published research and administrative work including committee membership (see 'How to get tenure'). Most tenure systems allow junior tenure-track faculty members a period of several years to establish such a record. In addition to job security, academic tenure aims to protect academic freedom; faculty members can disagree with popular opinion, express negative views about their institution, or research unpopular topics.

Nevertheless, tenure is receding in the United States, where tight budgets have prompted universities to hire more adjunct faculty members. In 1970, roughly three-quarters of all faculty members were in the tenure stream in the United States, according to figures amassed by the American Association of University Professors (AAUP). By 1975, that number had dropped to 56%, and it continued to fall. Only 42% received tenure in 1995, ►

IMAGES.COM/CORBIS

► and this had dropped to 30% by 2007. In the EU, tenure's availability varies widely depending on the state and institution, and it doesn't always confer the same benefits on scientists that it does in the United States.

Given the odds, is tenure still worth the struggle in today's competitive academic environment? It depends on what the researcher wants, says Marc Bousquet, an associate professor at Santa Clara University in California and a member of the executive council of the AAUP. "A lot of people think that tenure is the gold standard for job security, and it's often defined as lifetime job security," says Bousquet. "But tenure is a status marker," he says. "What the tenure system offers is a guarantee that the people doing the work are doing it at the highest possible level."

Tenure-track and tenured faculty members often become part of departmental governance and activities. They receive start-up packages, office space and financial support from the university; for example, they often get a bridge grant to tide them over until their first external funding award. Non-tenure-track faculty members in the United States and in Europe are generally ineligible for such grants from their institution and have to seek 'soft' money from individual research grants instead.

Critics of tenure in the United States and Europe say that it effectively bars junior faculty members from getting hired at institutions filled with senior tenured faculty members, and allows senior faculty members to become unproductive and complacent. Contingent faculty members — contract employees usually lacking benefits — lament the lack of financial and peer support and interaction typically associated with non-tenure-track positions. A May 2010 report on contingent faculty members by the Center for the Education of Women (CEW) at the University of Michigan in Ann Arbor found that many feel shunned by their tenured colleagues, often working in isolation and barred from departmental activities and governance. "They don't get the chance to see other research faculty," says Jean Waltman, senior associate at the CEW and lead author of the study, adding that there is often no institutionalized way to bring together contingent faculty members.



"If I secure funds to continue paying my own salary, I can conceivably stay here as long as I want."

Rafael Carazo Salas

Contingents in the CEW study also reported that they were at the mercy of their department chair. Those with good chair relationships found that they could function in much the same way as

a tenure-track faculty member — attending and speaking at department meetings, voting on governance issues and contributing to the university's science community. Poor chair relationships, however, meant that the non-tenure-track researcher was apt to flounder and more likely to be told that his or her contract would not be renewed — whether there was financial cause or not.

At European institutions, academic researchers who work under fixed-term contracts face many of these same problems, says Marja Makarow, chief executive at the European Science Foundation (ESF), headquartered in Strasbourg, France. "We really lack clear career pathways in Europe for young investigators," says Makarow, who is a former vice-rector for research at the University of Helsinki. "Universities need to create attractive positions for this generation and assure them that there are viable career possibilities for them in the academic environment."



"We really lack clear career pathways in Europe for young investigators."

Marja Makarow

Tenure does not, however, provide a respite from the rigours of a researcher's duties — in some ways, it heightens them. "The pressure isn't over," says Matthew Ames, chair of the department of molecular pharmacology and experimental therapeutics at the Mayo Clinic in Rochester, Minnesota. "Having tenure has never made me sleep better at night." Non-tenure-track research faculty members who participated in the CEW study said that their tenured colleagues often face significant administrative pressures and obligations that stymie their research. "They are so down in the trenches, researching the nitty, nitty, nitty, nitty gritty," reported one respondent. "I don't have so many strings, and I'm not tied up in meetings and all those other obligations, so I have the time to think more creatively and to step back and take some risks."

AN ACADEMIC AND ECONOMIC ELITE

Tenure is also, in a sense, very expensive. In the United States, the increasingly lengthy time to tenure, during which scientists earn relatively low wages as graduate students and postdocs, means that tenure often becomes the province of the financially privileged, according to Bousquet. "The tenured have always been an academic elite, but they have not always been drawn only from our economic elite," he says. "For many decades, we made it possible for persons of middle, lower-

TIPS FROM ACADEMICS

How to get tenure

- Develop a strong, clear research profile. Link your name to your specific research focus by publishing often, attending conferences and speaking about your work, giving talks, creating poster presentations and providing synopses and discussion on your website. Also, line up two or three senior collaborators who can help to broadcast your work. Don't always collaborate with one particular person, as that may create the impression that you can't do the work on your own.
- Be a good citizen in your department. Portray yourself as a department member who gets along with all colleagues, who will advocate for other department members and make sure that things get done. Name a senior department member as a co-author on your papers if you're in Europe.
- Network. Contact several colleagues — whom you might have met at conferences or elsewhere — who are experts on your research topic. Develop a continuing exchange with them about your work. Such contacts may help to write letters to the tenure-review committee about your research.
- Develop two mentors. One, more senior than you, will help you to build your research programme and prioritize your work in accordance with your institution's tenure criteria. The other, more junior than the first mentor, can work with you 'on the ground' and help you to understand your institution's tenure requirements and expectations. For example, if you're in the fourth year of your tenure clock, you can't go back to teach a second undergraduate class if you thought you had to teach only one. This younger mentor can help you to keep these requirements straight.
- Teach. Yes, it takes away from your time at the bench. But in the United Kingdom, there are only two paths to positions that resemble tenure, and one requires teaching. In the United States, if you're not on the tenure track but want to be, teaching duties will help.
- Maintain confidence. Trust that you're going to make it. Keep focused on your research and don't lose sleep worrying. Anxiety and sleepless nights will do little but make you work less effectively. **K.K.**

middle and lower-class backgrounds to find their way into the professoriate.”

But it is not clear whether the other options are any better. Most alternative proposals call for some variation of the mid- to long-term contracts already in place at non-university institutions such as the Howard Hughes Medical Institute's Janelia Farm Research Campus in Ashburn, Virginia. Picking up on this trend, several universities have created similar arrangements, adopting provisions for improved job security or inclusion in governance. The California State University system, for example, offers fixed-term contracts to adjuncts and requires exploring alternatives to layoffs. The City University of New York offers eligibility for ‘continuous employment’ to contingent faculty members who have completed six years of service.

Some suggest that a contract system is better than existing European-style tenure. Natalie Sebanz, a tenured associate professor at Radboud University in Nijmegen, the Netherlands, says that earning tenure in the EU is often not transparent or straightforward, even for tenure-track faculty members who have done all the requisite work. “The decision is not always fully based on scientific merit,” says Sebanz. “You need to know the right people and say particular things and not say other things. It has a lot to do with the old existing network.” Nor does tenure in the EU always offer the same advantages that US researchers take for granted, such as scientific independence and autonomy. Junior faculty members at some institutions, even if they're on the tenure track, often have to defer to the department chair or a more senior faculty member for approval of their research topic or for the right to be an adviser to PhD students.

For those seeking autonomy and independence, a contract post is the way to go, agrees Moritz Daum, head of a research group on infant cognition at the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig, Germany. Daum has a six-year contract, which is standard at Planck institutes, none of which offers tenure to researchers. As a contract scientist, Daum says he defers to no one. “You don't have a director or professor telling you what to do,” he says.

Still, for many young scientists in any location, autonomy and independence can trump job security for only so long. The ESF report found that the missing combination of job security, good pay and mobility is a major reason for Europe's dwindling academic research workforce. To boost recruiting efforts, some universities in the EU — including the ETH, the University of Helsinki, and Aalto University in Helsinki and Espoo, Finland — are piloting the tenure-track concept. Despite the criticisms levelled against it, many consider US-style tenure to be superior



“Having tenure has never made me sleep better at night.”

Matthew Ames

to tenure and tenure track at most European institutions, if only because of its transparency, says Makarow. The motivation to make European institutions more attractive to researchers is more than just a general aspiration. It is, in part, an outgrowth of the European Commission's plan to boost the EU's investment in research and development to 3% of its annual gross domestic product by 2020 as part of an economic-growth strategy. To achieve that target, the commission has estimated an additional 1 million researchers would be needed in the next ten years — no paltry number.

It is not clear whether the EU will offer new models of tenure or whether the United States will produce acceptable, viable alternatives. But researchers can take heart from the fact that global change is afoot, slow though it may seem, says Beate Scholz, former chair of the European Science Foundation's member organization forum on research careers and lead author of the January 2010 report *Research Careers in Europe: Landscape and Horizons*. “This,” says Scholz, “is a system in transformation.”

Scholz says tenure may gradually appear more consistently throughout the EU within the next decade or so. Like Makarow, she also believes that the creation of transparent tenure-track and tenured positions at universities would be an effective recruitment tool, and calls for them in the ESF report. “A research career in Europe is very insecure. Researchers are not able to tell what the next step is, whether their contract will be renewed,” says Scholz. “Many people would rather drop out.”

Reports of tenure's demise — at least in some regions — may turn out to be greatly exaggerated. “What else besides extraordinary job security is going to attract someone willing to put in more than ten years of post-graduate study?” asks Bousquet. Despite tenure's drawbacks, Ames says that it has allowed him to conduct better research because he knows his job is secure. “I could have a vision and long-range plan, and build my research programme based on my understanding that I would be here the following year and the year after that and the year after that,” he says. “Research, to me, is not a one- to two-year cycle. It requires stability and continuity.” ■

Karen Kaplan is the assistant Careers editor at Nature.

EUROPE

PhD improvements

To improve the quality of Europe's academic research, its higher-education institutions must upgrade doctoral programmes, says the European University Association (EUA). Representing some 900 institutions, the EUA adopted a series of recommendations on 21 October. These include: developing a variety of research programmes in different fields at individual institutions; creating supervisory guidelines and training programmes for areas such as project management; and giving PhD researchers supervision, career support and the opportunity to develop their own research topics. “A good part of the research that goes on in Europe is done by doctoral students,” says Thomas Jørgensen, head of the EUA's council for doctoral education. “We cannot waste this resource.”

EQUAL OPPORTUNITIES

Transcending disability

Students with disabilities will learn how to conduct biomedical experiments independently, at a new Institute for Accessible Science at Purdue University in West Lafayette, Indiana. A \$2-million National Institutes of Health grant will support at least six students when it opens in 2012. Principal investigator Bradley Duerstock, who uses a wheelchair, says there aren't enough resources for young disabled researchers, and challenges such as reaching into a fume hood often discourage students from pursuing science. Online mentoring topics will include lab modifications and remotely accessing equipment. The grant is one of six NIH awards supporting diversity.

EUROPEAN UNION

Progress on pensions

The European Commission (EC) could decide on how to set up transferable pensions for researchers by 15 November, an EC spokesman says. Meanwhile, the EC will accept public comment on previously proposed changes to the pensions framework (see go.nature.com/nbqfcy). Solutions under consideration include the creation of a pan-European pension fund for researchers, a step that European legislation would allow, according to EC spokesman Mark English. “The commission would be prepared to do everything within its power to facilitate such a move,” he says.

middle and lower-class backgrounds to find their way into the professoriate.”

But it is not clear whether the other options are any better. Most alternative proposals call for some variation of the mid- to long-term contracts already in place at non-university institutions such as the Howard Hughes Medical Institute's Janelia Farm Research Campus in Ashburn, Virginia. Picking up on this trend, several universities have created similar arrangements, adopting provisions for improved job security or inclusion in governance. The California State University system, for example, offers fixed-term contracts to adjuncts and requires exploring alternatives to layoffs. The City University of New York offers eligibility for ‘continuous employment’ to contingent faculty members who have completed six years of service.

Some suggest that a contract system is better than existing European-style tenure. Natalie Sebanz, a tenured associate professor at Radboud University in Nijmegen, the Netherlands, says that earning tenure in the EU is often not transparent or straightforward, even for tenure-track faculty members who have done all the requisite work. “The decision is not always fully based on scientific merit,” says Sebanz. “You need to know the right people and say particular things and not say other things. It has a lot to do with the old existing network.” Nor does tenure in the EU always offer the same advantages that US researchers take for granted, such as scientific independence and autonomy. Junior faculty members at some institutions, even if they're on the tenure track, often have to defer to the department chair or a more senior faculty member for approval of their research topic or for the right to be an adviser to PhD students.

For those seeking autonomy and independence, a contract post is the way to go, agrees Moritz Daum, head of a research group on infant cognition at the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig, Germany. Daum has a six-year contract, which is standard at Planck institutes, none of which offers tenure to researchers. As a contract scientist, Daum says he defers to no one. “You don't have a director or professor telling you what to do,” he says.

Still, for many young scientists in any location, autonomy and independence can trump job security for only so long. The ESF report found that the missing combination of job security, good pay and mobility is a major reason for Europe's dwindling academic research workforce. To boost recruiting efforts, some universities in the EU — including the ETH, the University of Helsinki, and Aalto University in Helsinki and Espoo, Finland — are piloting the tenure-track concept. Despite the criticisms levelled against it, many consider US-style tenure to be superior



“Having tenure has never made me sleep better at night.”

Matthew Ames

ment in research and development to 3% of its annual gross domestic product by 2020 as part of an economic-growth strategy. To achieve that target, the commission has estimated an additional 1 million researchers would be needed in the next ten years — no paltry number.

It is not clear whether the EU will offer new models of tenure or whether the United States will produce acceptable, viable alternatives. But researchers can take heart from the fact that global change is afoot, slow though it may seem, says Beate Scholz, former chair of the European Science Foundation's member organization forum on research careers and lead author of the January 2010 report *Research Careers in Europe: Landscape and Horizons*. “This,” says Scholz, “is a system in transformation.”

Scholz says tenure may gradually appear more consistently throughout the EU within the next decade or so. Like Makarow, she also believes that the creation of transparent tenure-track and tenured positions at universities would be an effective recruitment tool, and calls for them in the ESF report. “A research career in Europe is very insecure. Researchers are not able to tell what the next step is, whether their contract will be renewed,” says Scholz. “Many people would rather drop out.”

Reports of tenure's demise — at least in some regions — may turn out to be greatly exaggerated. “What else besides extraordinary job security is going to attract someone willing to put in more than ten years of post-graduate study?” asks Bousquet. Despite tenure's drawbacks, Ames says that it has allowed him to conduct better research because he knows his job is secure. “I could have a vision and long-range plan, and build my research programme based on my understanding that I would be here the following year and the year after that and the year after that,” he says. “Research, to me, is not a one- to two-year cycle. It requires stability and continuity.” ■

Karen Kaplan is the assistant Careers editor at Nature.

to tenure and tenure track at most European institutions, if only because of its transparency, says Makarow. The motivation to make European institutions more attractive to researchers is more than just a general aspiration. It is, in part, an outgrowth of the European Commission's plan to boost the EU's invest-

EUROPE

PhD improvements

To improve the quality of Europe's academic research, its higher-education institutions must upgrade doctoral programmes, says the European University Association (EUA). Representing some 900 institutions, the EUA adopted a series of recommendations on 21 October. These include: developing a variety of research programmes in different fields at individual institutions; creating supervisory guidelines and training programmes for areas such as project management; and giving PhD researchers supervision, career support and the opportunity to develop their own research topics. “A good part of the research that goes on in Europe is done by doctoral students,” says Thomas Jørgensen, head of the EUA's council for doctoral education. “We cannot waste this resource.”

EQUAL OPPORTUNITIES

Transcending disability

Students with disabilities will learn how to conduct biomedical experiments independently, at a new Institute for Accessible Science at Purdue University in West Lafayette, Indiana. A \$2-million National Institutes of Health grant will support at least six students when it opens in 2012. Principal investigator Bradley Duerstock, who uses a wheelchair, says there aren't enough resources for young disabled researchers, and challenges such as reaching into a fume hood often discourage students from pursuing science. Online mentoring topics will include lab modifications and remotely accessing equipment. The grant is one of six NIH awards supporting diversity.

EUROPEAN UNION

Progress on pensions

The European Commission (EC) could decide on how to set up transferable pensions for researchers by 15 November, an EC spokesman says. Meanwhile, the EC will accept public comment on previously proposed changes to the pensions framework (see go.nature.com/nbqfcy). Solutions under consideration include the creation of a pan-European pension fund for researchers, a step that European legislation would allow, according to EC spokesman Mark English. “The commission would be prepared to do everything within its power to facilitate such a move,” he says.

BY THE NUMBERS

Switzerland

BY KATHARINE SANDERSON

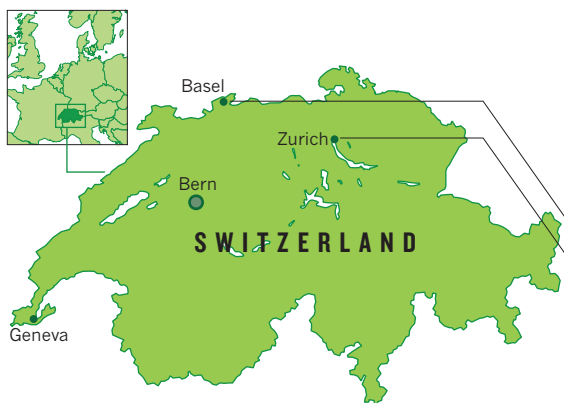
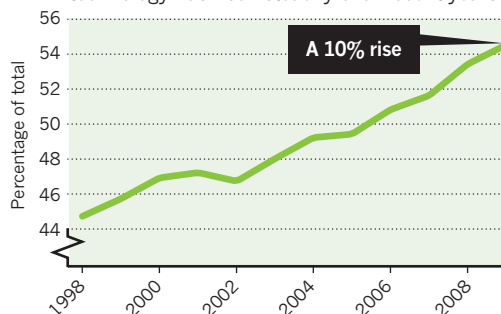
Boasting the world's most competitive economy — on the basis of factors such as business climate, institutions, infrastructure, higher education and innovation, according to the non-profit World Economic Forum — Switzerland also hosts some of Europe's finest universities, such as the ETH Zurich and the Ecole Polytechnique Fédérale de Lausanne. It also maintains a thriving pharmaceutical and biotech industry, along with impressive mountain scenery, fine watches and high-quality chocolate. Switzerland hosts the headquarters of several major drug-makers. The Basel area alone is home to the bases of Novartis, Roche and Actelion, and Geneva hosts headquarters for the pharmaceutical company Merck Serono.

BIOTECH AND PHARMA

- Some **245** pharmaceutical companies operate in Switzerland, employing **35,400** people, including researchers.
- Revenues from Swiss biotech companies rose from **8.09 billion** Swiss francs (US\$8.37 billion) in 2007 to **9.34 billion** Swiss francs in 2009, an increase of **15.4%**.
- The number of biotechs in Switzerland rose from **119** in 2001 to **162** in 2009. Biotech employees totalled **19,071** in 2009. Between 20% and 30% of those are researchers.

EMPLOYMENT IN SWITZERLAND

The number of employees in science and technology has risen steadily over recent years.



NANOTECHNOLOGY SWISS NANOSCIENCE INSTITUTE

- Based in Basel.
- Set up by the Swiss National Science Foundation.
- Produced **23** patents and **7** start-up companies.

IBM EUROPEAN RESEARCH BRANCH

- Based in Rüschlikon, Zurich.
- Building a 6,000-square-metre nanotechnology research lab, in collaboration with ETH Zurich, due to begin operations May 2011.

RESEARCH AND DEVELOPMENT (R&D) EXPENDITURES

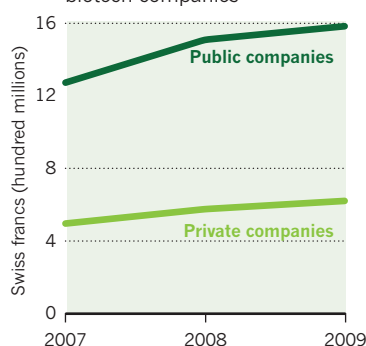
- Around **3%** of Switzerland's gross domestic product goes to R&D. Private industry accounts for **73%** of that amount.
- The Swiss government, industry and academia collectively spent **16.3 billion** Swiss francs on R&D in 2008, up **24%** since 2004.

INTERNATIONAL RESEARCHERS

- More than **50%** of professors hired between 2005 and 2008 in Switzerland's **12** research universities are from outside Switzerland.
- 45.7%** of professors and **50.3%** of assistants and scientific personnel are non-Swiss.
- Up to **56%** of professors are predicted to be non-Swiss by 2018.

SWISS R&D IN BIOTECH

R&D expenditure in Swiss biotech companies



Q&A Ralph Eichler



Ralph Eichler is president of ETH Zurich in Switzerland, which was 15th in the Times Higher Education's recent top world universities list — the highest ranked non-US and non-UK institution.

Why are half of Switzerland's researchers from other countries?

There is an international tradition. As we are a small country we have to search for the best talent worldwide. We also fund high-risk projects. The quality of life is an added value. There is low crime, political stability, a good education system and the security that comes with a good economy. People who come to work in Switzerland tend to stay here.

How does the ETH create successful spin-out companies?

A London School of Economics study found that 88% of ETH-founded companies survive more than five years. The ETH has a programme that links students to experienced business people to help them create a business plan. The students stay in the university for two years after graduating, setting up and running companies, and then have to become independent. This process, which is about 12 years old, clearly works.

What career advice would you give young Swiss scientists?

Whether in industry or academia, you should go abroad. About 80% of Swiss industry is geared towards exports, so we have to know the culture of different countries. I would never hire a professor who has not worked outside Switzerland. Good ETH researchers should go away and prove themselves in a second, different environment.

Do you have any trouble recruiting?

The biggest difficulty is the 'dual-career' issue. The ETH offers a dual-career programme, in which the partners of prospective employees are given support to find work. But it can still be a problem. **K.S.**

BUSH MEAT

A slick operation.

BY PAUL RENAULT

Was it honey and spoiled milk, or the deserts of love? Ray kneaded her with the stuff to keep the skin supple and elastic, to prevent tearing, to minimize the risk of infection and the amount of scarring afterwards. Part of him hated the scalpel, its forced entry.

Instruments and vinyl-gloved hands thrust into her plumbing — and there it was: the blockage. He could feel it, while they seemed more insensitive and detached than usual. Distracted. Incredulous. They argued with him; said he was seeing things, asked if he'd smoked 'more of that good Congolese ganja' prior to the procedure.

"I don't care if it is legal in your crackpot nation, or if you aren't technically under the influence here, when you're on the job. It's policy. Understand? You fail another drug test, Ray, we're going to have to let you go. Be a shame, losing a good spongebob like you. Might have to close up shop here, if we don't find a qualified replacement. We pull the plug, where does that leave you? Where does that leave *her* and the others? It's your choice..."

He took a deep breath and slowly let it out. He held his tongue. He had to stick one of the spare camera-wands in there manually and point it out to them before they addressed the problem in language long dulled by routine, in voices hollowed out like drinking gourds on their trip through space — via satellite — from halfway around the world. No apologies, no excuses, never a show of appreciation, nothing. The octopoid mechanical arms darted into and out of the patient with choreographed whirring precision. Ray turned to the screen mounted high on the wall and watched, keeping his thoughts to himself.

They left behind a trail of orange 'road-hazard' antiseptic and closed her up with surgical glue, even though he'd asked for stitches, worried that the glue would not hold. Ray had recurring nightmares of patients coming unglued, and there were few cases of malpractice here — a grieving family member was more likely to come after him with a machete, hack him to pieces, while the surgeons themselves would be safely out of reach.

But there was no discussing the issue. They had no time to spare.

➔ **NATURE.COM**
Discuss this story
online at:
go.nature.com/stcgua

Their surgical theatre included most of sub-Saharan Africa, parts of south and southeast

Asia, and many island nations in the Pacific. They flitted here and there like hummingbirds in a garden of the flesh. They left him to bathe and dress the patient. He switched her cart over to battery power, wheeled her into the post-op room for observation and plugged her back into the grid. Their cameras and bugs only made him feel that much more isolated. He tried not look at the patient, that lump under a white sheet with only the head sticking out. He turned away from her and headed to the locker room to strip off his scrubs and wash his hands at the sink. He stood at the mirror and considered a shave, left the razor on the shelf and turned back as though connected to the patient by an umbilical cord; an astronaut out on a spacewalk, wanting only to be cast adrift — to be able to look down on Earth and know that he was free at last.

He ran his time card at the desk and stepped outside to lock up for the night. The air felt — it felt like the inside of her body. Ray crossed the red dusty road to the café, where he met Gabon Harry at a table in the back. Harry was a big-boned man with a way of talking that picked you up and carried you along like floodwaters, often into the kind of trouble that you had to fight your way out of, but his years as a peacekeeper with the AU, including tours in Darfur and Mogadishu,

had mellowed him; so had the banana beer and the ganja. He blew smoke at the ceiling fan.

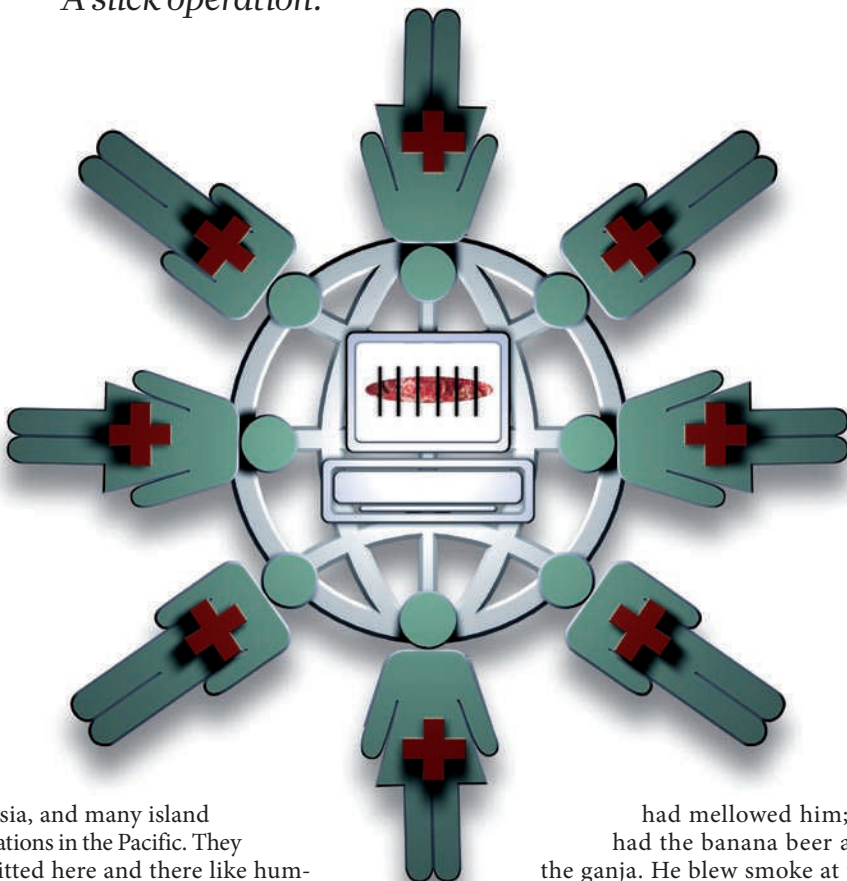
"You look beat, Ray. Why don't you quit that spongebobbing and come work for me? Take a load off, maybe live to a ripe old age? I can't afford to pay much now but once the maglev train comes through, we'll be going places. Why don't you come along?"

"I'll think about it."

"What's to think about?"

Her body on a slab in the building across the road. Her vitals up- and downloaded to a bank of computers, where GRCC nursing students drank energy drinks and chit-chatted to stay awake as they watched over patients reduced to figures on the screen. They learned to pick up and examine the torn and bloodied fragments of our love as though pawing through the bush meat in a cooler. We aren't real to them. We aren't people. We're broken things to be tinkered with, maybe to be fixed. Ray envied them, in a way — they burgled haunted houses and did not believe in ghosts; did not hear the screams; would not come undone with the glue. ■

Paul Renault took his first steps in Angola, picked up a second-hand language and kept on walking. He served time in (of all places) Grand Rapids, Michigan and Montreal.



JACEY

Moulting tail feathers in a juvenile oviraptorosaur

ARISING FROM Xing Xu, Xiaoting Zheng & Hailu You *Nature* **464**, 1338–1341 (2010)

Xu *et al.*¹ describe the extraordinarily preserved feathers from two subadults of the oviraptorosaur *Similicaudipteryx* from the Yixian Formation of Liaoning, China. The preserved tail feathers of the juvenile specimen (STM4.1) show a morphology not previously observed in any fossil feathers. The tail feathers of an older, immature specimen (STM22-6) show a typical closed pennaceous structure with a prominent, planar vane. I propose that the feathers of the tail of the juvenile specimen are not a specialized feather generation, but fossilized ‘pin feathers’ or developing feather germs.

Xu *et al.*¹ interpret the juvenile *Similicaudipteryx* tail feathers as examples of “proximately ribbon-like pennaceous feathers” (PRPFs) that have evolved convergently in avialan confuciusornithids and enanthiornithines, and in the non-avian maniraptoran *Epidexipteryx*. They describe the differences between juvenile and immature *Similicaudipteryx* feathers as a notable example of post-nestling ontogenetic change in feather morphology, and claim that “this phenomenon is not known to occur in other birds.”

Although modern birds do not show radical changes in flight-feather morphology after the nestling stage (probably owing to the functional constraints of flight), there are many examples of radical post-nestling changes in the morphology of other feathers. For example, the feather follicles on the head of a Wild Turkey (*Meleagris gallopavo*) grow plumulaceous down in nestlings, fully pennaceous contour feathers in juveniles, and specialized bristles in adults².

Data from Xu *et al.*¹ document that the tail feathers of the juvenile *Similicaudipteryx* specimen are very different in morphology from previously described PRPFs, in which the ribbon-like basal portion is formed completely by the lateral expansion and vertical compression of the rachis (Supplementary Fig. 3 in Xu *et al.*¹). In contrast, the unbranched, proximal portion of the tail feathers of the juvenile *Similicaudipteryx* is not a merely a continuous, basal extension of the rachis (Supplementary Fig. 2a in Xu *et al.*¹). Rather, the undifferentiated basal portion of these feathers surrounds the entire base of the exposed distal vane of the feather. The distal rachis does not expand laterally, and is visibly distinct from the broad base of the feather (top of Supplementary Fig. 2a in Xu *et al.*¹). This observation is inconsistent with the conclusion of Xu *et al.*¹

The juvenile tail feathers of *Similicaudipteryx* (STM4.1) are entirely consistent with the morphology of moulting feathers of living birds³. The tubular feather germ, called a pin feather, is surrounded by a keratinized sheath that falls off to expose the mature feather. Typically, the sheath begins to fall off the mature distal tip of the feather before the development of the ensheathed tubular base of the feather is complete. For a pennaceous feather, this intermediate stage of growth appears as a limited distal vane emerging from a smooth tubular base² (Fig. 1). These feathers are not too large to be ensheathed because feathers need to grow to their final size and may shed the sheath at different lengths. Indeed, the moulting Great Horned Owl feathers in Fig. 1 are over 6-cm long, and are mostly ensheathed. These structures are also not too wide to be ensheathed feathers; the width of pin feathers depends on the number of barbs, the size of barb ridges and the angle of barb-ridge expansion after emergence³. These parameters all vary among living birds. Tiny natal down feathers sometimes adhere to the distal tip of molting feathers in living birds, but this is rare after the nestling stage and would be unexpected in juvenile *Similicaudipteryx*.

I propose that the tail feathers of the juvenile *Similicaudipteryx* are similar, or identical, in morphology to those of the immature



Figure 1 | Developing primary wing feathers of a nestling Great Horned Owl. The distal tip of the planar vane of the pennaceous feathers are emerging from the tubular feather sheath. The sheath surrounds the entire base of the emergent vane, and the rachis runs under the sheath without lateral expansion. Photograph reproduced with permission from B. Hilton.

Similicaudipteryx, but in the juvenile specimen they were preserved at an intermediate stage of development during emergence from the distal portion of the vane from the feather sheath. Little is known about the evolutionary origin of feather moult, and the discovery of fossil moulting feathers is an extraordinary achievement in dinosaur palaeontology. The title of Xu *et al.*¹—“Exceptional dinosaur fossils show ontogenetic development of early feathers”—is still entirely consistent with this interpretation, but “ontogenetic” would refer to the growth of individual feathers rather than change over life.

The apparent simultaneous moult of tail feathers in *Similicaudipteryx* is similar to the first set of flight feathers in juvenile modern birds, but differs from adult modern birds in which tail feathers moult sequentially. However, sequential moult of remiges and rectrices is proposed to have evolved as an adaptation for flight to maintain aerodynamic function. Early feathered theropods would be unlikely to have evolved sequential moult. Simultaneous rectrix molt in terrestrial theropods may have preceded the evolution of sequential rectrix moult in modern birds.

Richard O. Prum¹

¹Department of Ecology and Evolutionary Biology, and Peabody Natural History Museum, Yale University, New Haven, Connecticut 06520-8150, USA.

e-mail: richard.prum@Yale.edu

Received 13 May; accepted 18 August 2010.

1. Xu, X., Zheng, X. & You, H. Exceptional dinosaur fossils show ontogenetic development of early feathers. *Nature* **464**, 1338–1341 (2010).
2. Lucas, A. M. & Stettenheim, P. R. *Avian Anatomy: Integument* (US Dept of Agriculture, 1972).
3. Prum, R. O. & Williamson, S. Theory of the growth and evolution of feather shape. *J. Exp. Zool.* **291**, 30–57 (2001).

Competing financial interests: declared none.

doi:10.1038/nature09480

Xu *et al.* replyREPLYING TO Richard O. Prum *Nature* **468**, doi:10.1038/nature09480

Prum¹ identifies the unusual feathers in the smaller *Similicaudipteryx* specimen as immature, and suggests that the different morphologies preserved in the smaller and larger specimens represent different growth stages of a single feather type rather than successive feather generations of different types. Although this new proposal is very interesting, we do not agree that it is better supported by the available data than our original interpretation².

A feather of any generation can be expected to show different morphologies in the course of its growth³, as noted by Prum¹. During early growth, the feather will be enclosed in a long sheath. Subsequently, the sheath will begin to open from the distal end, revealing the pennaceous feather. For simplicity, we refer to feathers at this early stage of opening as emergent feathers. The proximally ribbon-like pennaceous feathers (PRPFs) in the *Similicaudipteryx* specimen are superficially similar to the emergent feathers of extant birds³, but they are apparently different in relative size. Because feather growth is still far from complete when the sheath begins to open, emergent feathers are proportionally small³, and this is particularly true in early feather generations. The overall immaturity of the smaller *Similicaudipteryx* specimen implies that its plumage would have been of an early generation², so that emergent feathers of large size would be particularly surprising in this individual. However, the PRPFs of the smaller *Similicaudipteryx* are longer than the tibiotarsus, greatly exceeding the proportional length of early-generation emergent feathers in modern birds.

Also, the ribbon-like portion of a typical PRPF in the smaller *Similicaudipteryx* specimen is only slightly narrower than the pennaceous portion, and its width considerably exceeds the length of the associated caudal vertebra. In contrast, the ensheathed portion of an emergent feather of an extant bird is much narrower than the pennaceous portion. Also, the identification of all rectrices and remiges in the *Similicaudipteryx* specimen as emergent feathers would imply that *Similicaudipteryx* moulted all of its flight feathers simultaneously, in contrast to the sequential moulting of extant birds¹. The interpretation by Prum¹ of simultaneous moulting as a primitive condition present in non-avian dinosaurs is interesting, but at face value this is nevertheless an obstacle to interpreting the plumage of the smaller *Similicaudipteryx* specimen as a generation of emergent feathers. Finally, because new feathers physically displace old feathers during the process of moulting, a displaced mature feather will often be found hanging from the tip of the emergent feather that is replacing it. The total absence of previous-generation feathers associated with the PRPFs of *Similicaudipteryx* represents another piece of evidence indicating that the PRPFs are unlikely to represent emergent feathers, unless the replacement process was substantially different from that seen in extant birds.

Prum¹ emphasizes some differences between the PRPFs of the *Similicaudipteryx* specimen and those of basal birds and the non-avian maniraptoran *Epidexipteryx*. We noted these differences in our original study², but Prum's alternative interpretation led us to reconsider the nature of the PRPFs in the *Similicaudipteryx* specimen and those in basal birds and *Epidexipteryx*². The ribbon-like portion of the PRPFs in basal birds and *Epidexipteryx* appears to represent an extremely broad rachis, given that the definite rachis of the pennaceous portion is identifiable as a tapering extension of the entire ribbon-like portion⁴. Under this interpretation, the median line that typically extends along the ribbon-like portion of the feather, and which was previously identified as a rachis⁵, would actually represent a ventral furrow running along the rachis as in modern birds.

In contrast to both the broad proximal rachises of basal birds and *Epidexipteryx*, and the horny feather sheaths referred to by Prum¹, the ribbon-like portions of the PRPFs of the smaller *Similicaudipteryx* specimen may represent hollow structures similar to the broad monofilaments seen in more basal theropods⁶. Admittedly, two-dimensional preservation makes it difficult to judge which identification is correct, but further early juvenile *Similicaudipteryx* specimens with normal flight feathers would support Prum's interpretation. Nevertheless, differences in feather ontogeny between Mesozoic and extant taxa indicate that analogies involving specific feather structures should be drawn only with extreme caution.

Xing Xu¹, Xiaoting Zheng² & Hailu You³

¹Key Laboratory of Evolutionary Systematics of Vertebrates, Institute of Vertebrate Paleontology & Paleoanthropology, Chinese Academy of Sciences, 142 Xiwai Street, Beijing 100044, China.

e-mail: xingxu@vip.sina.com

²Shandong Tianyu Museum of Nature, Pingyi, Shandong 273300, China.

³Institute of Geology, Chinese Academy of Geological Sciences, 26 Baiwanzhuang Road, Beijing 100037, China.

1. Prum, R. O. Moulting tail feathers in a juvenile oviraptorosaur. *Nature* **468**, doi:10.1038/nature09480 (2010).
2. Xu, X., Zheng, X. & You, H. Exceptional dinosaur fossils show ontogenetic development of early feathers. *Nature* **464**, 1338–1341 (2010).
3. Lucas, A. M. & Stettenheim, P. R. *Avian Anatomy: Integument* (US Dept of Agriculture, 1972).
4. Zheng, X. T. *The Origin of Birds* 55–58 (Shandong Science and Technology Press, 2009).
5. Zhang, F., Zhou, Z., Xu, X., Wang, X. & Sullivan, C. A bizarre Jurassic maniraptoran from China with elongate ribbon-like feathers. *Nature* **455**, 1105–1108 (2008).
6. Xu, X., Zheng, X. & You, H. A new feather type in a nonavian theropod and the early evolution of feathers. *Proc. Natl Acad. Sci. USA* **106**, 832–834 (2009).

Competing financial interests: declared none.

doi:10.1038/nature09481

Moulting tail feathers in a juvenile oviraptorosaur

ARISING FROM Xing Xu, Xiaoting Zheng & Hailu You *Nature* **464**, 1338–1341 (2010)

Xu *et al.*¹ describe the extraordinarily preserved feathers from two subadults of the oviraptorosaur *Similicaudipteryx* from the Yixian Formation of Liaoning, China. The preserved tail feathers of the juvenile specimen (STM4.1) show a morphology not previously observed in any fossil feathers. The tail feathers of an older, immature specimen (STM22-6) show a typical closed pennaceous structure with a prominent, planar vane. I propose that the feathers of the tail of the juvenile specimen are not a specialized feather generation, but fossilized ‘pin feathers’ or developing feather germs.

Xu *et al.*¹ interpret the juvenile *Similicaudipteryx* tail feathers as examples of “proximately ribbon-like pennaceous feathers” (PRPFs) that have evolved convergently in avialan confuciusornithids and enanthiornithines, and in the non-avian maniraptoran *Epidexipteryx*. They describe the differences between juvenile and immature *Similicaudipteryx* feathers as a notable example of post-nestling ontogenetic change in feather morphology, and claim that “this phenomenon is not known to occur in other birds.”

Although modern birds do not show radical changes in flight-feather morphology after the nestling stage (probably owing to the functional constraints of flight), there are many examples of radical post-nestling changes in the morphology of other feathers. For example, the feather follicles on the head of a Wild Turkey (*Meleagris gallopavo*) grow plumulaceous down in nestlings, fully pennaceous contour feathers in juveniles, and specialized bristles in adults².

Data from Xu *et al.*¹ document that the tail feathers of the juvenile *Similicaudipteryx* specimen are very different in morphology from previously described PRPFs, in which the ribbon-like basal portion is formed completely by the lateral expansion and vertical compression of the rachis (Supplementary Fig. 3 in Xu *et al.*¹). In contrast, the unbranched, proximal portion of the tail feathers of the juvenile *Similicaudipteryx* is not a merely a continuous, basal extension of the rachis (Supplementary Fig. 2a in Xu *et al.*¹). Rather, the undifferentiated basal portion of these feathers surrounds the entire base of the exposed distal vane of the feather. The distal rachis does not expand laterally, and is visibly distinct from the broad base of the feather (top of Supplementary Fig. 2a in Xu *et al.*¹). This observation is inconsistent with the conclusion of Xu *et al.*¹

The juvenile tail feathers of *Similicaudipteryx* (STM4.1) are entirely consistent with the morphology of moulting feathers of living birds³. The tubular feather germ, called a pin feather, is surrounded by a keratinized sheath that falls off to expose the mature feather. Typically, the sheath begins to fall off the mature distal tip of the feather before the development of the ensheathed tubular base of the feather is complete. For a pennaceous feather, this intermediate stage of growth appears as a limited distal vane emerging from a smooth tubular base² (Fig. 1). These feathers are not too large to be ensheathed because feathers need to grow to their final size and may shed the sheath at different lengths. Indeed, the moulting Great Horned Owl feathers in Fig. 1 are over 6-cm long, and are mostly ensheathed. These structures are also not too wide to be ensheathed feathers; the width of pin feathers depends on the number of barbs, the size of barb ridges and the angle of barb-ridge expansion after emergence³. These parameters all vary among living birds. Tiny natal down feathers sometimes adhere to the distal tip of molting feathers in living birds, but this is rare after the nestling stage and would be unexpected in juvenile *Similicaudipteryx*.

I propose that the tail feathers of the juvenile *Similicaudipteryx* are similar, or identical, in morphology to those of the immature



Figure 1 | Developing primary wing feathers of a nestling Great Horned Owl. The distal tip of the planar vane of the pennaceous feathers are emerging from the tubular feather sheath. The sheath surrounds the entire base of the emergent vane, and the rachis runs under the sheath without lateral expansion. Photograph reproduced with permission from B. Hilton.

Similicaudipteryx, but in the juvenile specimen they were preserved at an intermediate stage of development during emergence from the distal portion of the vane from the feather sheath. Little is known about the evolutionary origin of feather moult, and the discovery of fossil moulting feathers is an extraordinary achievement in dinosaur palaeontology. The title of Xu *et al.*¹—“Exceptional dinosaur fossils show ontogenetic development of early feathers”—is still entirely consistent with this interpretation, but “ontogenetic” would refer to the growth of individual feathers rather than change over life.

The apparent simultaneous moult of tail feathers in *Similicaudipteryx* is similar to the first set of flight feathers in juvenile modern birds, but differs from adult modern birds in which tail feathers moult sequentially. However, sequential moult of remiges and rectrices is proposed to have evolved as an adaptation for flight to maintain aerodynamic function. Early feathered theropods would be unlikely to have evolved sequential moult. Simultaneous rectrix moult in terrestrial theropods may have preceded the evolution of sequential rectrix moult in modern birds.

Richard O. Prum¹

¹Department of Ecology and Evolutionary Biology, and Peabody Natural History Museum, Yale University, New Haven, Connecticut 06520-8150, USA.

e-mail: richard.prum@Yale.edu

Received 13 May; accepted 18 August 2010.

1. Xu, X., Zheng, X. & You, H. Exceptional dinosaur fossils show ontogenetic development of early feathers. *Nature* **464**, 1338–1341 (2010).
2. Lucas, A. M. & Stettenheim, P. R. *Avian Anatomy: Integument* (US Dept of Agriculture, 1972).
3. Prum, R. O. & Williamson, S. Theory of the growth and evolution of feather shape. *J. Exp. Zool.* **291**, 30–57 (2001).

Competing financial interests: declared none.

doi:10.1038/nature09480

Xu *et al.* replyREPLYING TO Richard O. Prum *Nature* **468**, doi:10.1038/nature09480

Prum¹ identifies the unusual feathers in the smaller *Similicaudipteryx* specimen as immature, and suggests that the different morphologies preserved in the smaller and larger specimens represent different growth stages of a single feather type rather than successive feather generations of different types. Although this new proposal is very interesting, we do not agree that it is better supported by the available data than our original interpretation².

A feather of any generation can be expected to show different morphologies in the course of its growth³, as noted by Prum¹. During early growth, the feather will be enclosed in a long sheath. Subsequently, the sheath will begin to open from the distal end, revealing the pennaceous feather. For simplicity, we refer to feathers at this early stage of opening as emergent feathers. The proximally ribbon-like pennaceous feathers (PRPFs) in the *Similicaudipteryx* specimen are superficially similar to the emergent feathers of extant birds³, but they are apparently different in relative size. Because feather growth is still far from complete when the sheath begins to open, emergent feathers are proportionally small³, and this is particularly true in early feather generations. The overall immaturity of the smaller *Similicaudipteryx* specimen implies that its plumage would have been of an early generation², so that emergent feathers of large size would be particularly surprising in this individual. However, the PRPFs of the smaller *Similicaudipteryx* are longer than the tibiotarsus, greatly exceeding the proportional length of early-generation emergent feathers in modern birds.

Also, the ribbon-like portion of a typical PRPF in the smaller *Similicaudipteryx* specimen is only slightly narrower than the pennaceous portion, and its width considerably exceeds the length of the associated caudal vertebra. In contrast, the ensheathed portion of an emergent feather of an extant bird is much narrower than the pennaceous portion. Also, the identification of all rectrices and remiges in the *Similicaudipteryx* specimen as emergent feathers would imply that *Similicaudipteryx* moulted all of its flight feathers simultaneously, in contrast to the sequential moulting of extant birds¹. The interpretation by Prum¹ of simultaneous moulting as a primitive condition present in non-avian dinosaurs is interesting, but at face value this is nevertheless an obstacle to interpreting the plumage of the smaller *Similicaudipteryx* specimen as a generation of emergent feathers. Finally, because new feathers physically displace old feathers during the process of moulting, a displaced mature feather will often be found hanging from the tip of the emergent feather that is replacing it. The total absence of previous-generation feathers associated with the PRPFs of *Similicaudipteryx* represents another piece of evidence indicating that the PRPFs are unlikely to represent emergent feathers, unless the replacement process was substantially different from that seen in extant birds.

Prum¹ emphasizes some differences between the PRPFs of the *Similicaudipteryx* specimen and those of basal birds and the non-avian maniraptoran *Epidexipteryx*. We noted these differences in our original study², but Prum's alternative interpretation led us to reconsider the nature of the PRPFs in the *Similicaudipteryx* specimen and those in basal birds and *Epidexipteryx*². The ribbon-like portion of the PRPFs in basal birds and *Epidexipteryx* appears to represent an extremely broad rachis, given that the definite rachis of the pennaceous portion is identifiable as a tapering extension of the entire ribbon-like portion⁴. Under this interpretation, the median line that typically extends along the ribbon-like portion of the feather, and which was previously identified as a rachis⁵, would actually represent a ventral furrow running along the rachis as in modern birds.

In contrast to both the broad proximal rachises of basal birds and *Epidexipteryx*, and the horny feather sheaths referred to by Prum¹, the ribbon-like portions of the PRPFs of the smaller *Similicaudipteryx* specimen may represent hollow structures similar to the broad monofilaments seen in more basal theropods⁶. Admittedly, two-dimensional preservation makes it difficult to judge which identification is correct, but further early juvenile *Similicaudipteryx* specimens with normal flight feathers would support Prum's interpretation. Nevertheless, differences in feather ontogeny between Mesozoic and extant taxa indicate that analogies involving specific feather structures should be drawn only with extreme caution.

Xing Xu¹, Xiaoting Zheng² & Hailu You³

¹Key Laboratory of Evolutionary Systematics of Vertebrates, Institute of Vertebrate Paleontology & Paleoanthropology, Chinese Academy of Sciences, 142 Xiwai Street, Beijing 100044, China.

e-mail: xingxu@vip.sina.com

²Shandong Tianyu Museum of Nature, Pingyi, Shandong 273300, China.

³Institute of Geology, Chinese Academy of Geological Sciences, 26 Baiwanzhuang Road, Beijing 100037, China.

1. Prum, R. O. Moulting tail feathers in a juvenile oviraptorosaur. *Nature* **468**, doi:10.1038/nature09480 (2010).
2. Xu, X., Zheng, X. & You, H. Exceptional dinosaur fossils show ontogenetic development of early feathers. *Nature* **464**, 1338–1341 (2010).
3. Lucas, A. M. & Stettenheim, P. R. *Avian Anatomy: Integument* (US Dept of Agriculture, 1972).
4. Zheng, X. T. *The Origin of Birds* 55–58 (Shandong Science and Technology Press, 2009).
5. Zhang, F., Zhou, Z., Xu, X., Wang, X. & Sullivan, C. A bizarre Jurassic maniraptoran from China with elongate ribbon-like feathers. *Nature* **455**, 1105–1108 (2008).
6. Xu, X., Zheng, X. & You, H. A new feather type in a nonavian theropod and the early evolution of feathers. *Proc. Natl Acad. Sci. USA* **106**, 832–834 (2009).

Competing financial interests: declared none.

doi:10.1038/nature09481

The amino-terminal disease hotspot of ryanodine receptors forms a cytoplasmic vestibule

Ching-Chieh Tung¹, Paolo A. Lobo¹, Lynn Kimlicka¹ & Filip Van Petegem¹

Many physiological events require transient increases in cytosolic Ca^{2+} concentrations. Ryanodine receptors (RyRs) are ion channels that govern the release of Ca^{2+} from the endoplasmic and sarcoplasmic reticulum¹. Mutations in RyRs can lead to severe genetic conditions that affect both cardiac and skeletal muscle, but locating the mutated residues in the full-length channel structure has been difficult^{2,3}. Here we show the 2.5 Å resolution crystal structure of a region spanning three domains of RyR type 1 (RyR1), encompassing amino acid residues 1–559. The domains interact with each other through a predominantly hydrophilic interface. Docking in RyR1 electron microscopy maps^{4,5} unambiguously places the domains in the cytoplasmic portion of the channel, forming a 240-kDa cytoplasmic vestibule around the four-fold symmetry axis. We pinpoint the exact locations of more than 50 disease-associated mutations in full-length RyR1 and RyR2. The mutations can be classified into three groups: those that destabilize the interfaces between the three amino-terminal domains, disturb the folding of individual domains or affect one of six interfaces with other parts of the receptor. We propose a model whereby the opening of a RyR coincides with allosterically coupled motions within the N-terminal domains. This process can be affected by mutations that target various interfaces within and across subunits. The crystal structure provides a framework to understand the many disease-associated mutations in RyRs that have been studied using functional methods, and will be useful for developing new strategies to modulate RyR function in disease states.

Ryanodine receptors are large (~2.2 MDa) tetrameric channels that control the release of Ca^{2+} from the endoplasmic and sarcoplasmic reticulum. Three isoforms (RyR1, RyR2 and RyR3) have been isolated in mammalian organisms. Excitation–contraction coupling primarily relies on RyR1 in skeletal muscle and RyR2 in cardiac muscle¹. Because Ca^{2+} is a potent messenger, the mishandling of Ca^{2+} release owing to point mutations in either isoform can lead to severe genetic diseases, including malignant hyperthermia, central core disease and multi-minicore disease for RyR1, and heart disorders such as catecholaminergic polymorphic ventricular tachycardia and arrhythmogenic right ventricular dysplasia for RyR2 (ref. 2). With a few exceptions, the point substitutions cluster in three distinct mutation ‘hotspots’³. RyRs are activated by Ca^{2+} , and most disease-associated mutations that have been analysed so far introduce a gain of function, either increasing the sensitivity of the RyR to activators or producing a leaky channel, or both⁶. Owing to the limited availability of high-resolution structures, a mechanistic description of the effect of the disease-associated mutations has been lacking.

To gain further insight into RyR function in physiological and disease states, we solved the 2.5 Å crystal structure of a 62-kDa region of rabbit RyR1 (amino acid residues 1–559) that covers most of the N-terminal hotspot (Fig. 1 and Supplementary Fig. 1). The structure was refined to a final R_{cryst} and R_{free} of 20.9% and 23.4%, respectively. The construct folds into three separate domains (A, B and C) that interact with each other through a mainly hydrophilic interface (Supplementary Table 1). Domain A (amino acid residues 1–205) forms a β -trefoil domain with

an additional α -helix, consistent with previous findings^{7,8}. Domain B (residues 206–394) forms another β -trefoil domain, whereas domain C (residues 395–532) forms a bundle of five α -helices. The carboxy-terminal 27 residues are not visible in the electron density and might be part of another domain. The domain interfaces cover a surface of 1,687 Å², of which 517 Å² is hydrophobic (Supplementary Table 1).

Cryo-electron microscopy (cryo-EM) studies have shown full-length RyR1 images at ~10 Å resolution^{4,5,9}. The availability of a crystal structure with more than 22,000 Å² of surface area enabled us to locate the 62-kDa N-terminal region unambiguously in the full-length RyR1 structure^{10,11}. The N-terminal domains are located in the cytoplasmic portion of the channel, forming a vestibule around the four-fold axis (Fig. 2, Supplementary Fig. 2 and Supplementary Movie 1).

To prove this result, we split the crystal structure into two parts (domain A and domain BC) and performed independent docking experiments with each part. Both parts dock to the same site and recapitulate the A–BC interface with near perfection (Fig. 2c). Given the vast number of possible orientations and locations of individual domains in a large cryo-EM map, the probability of reconstituting the crystal structure through chance is infinitesimally small (see Supplementary Discussion). In addition, the same position of the N-terminal domains is obtained with a very high docking contrast using three RyR1 maps of different resolution (9.6 Å, 10.3 Å and 14 Å)^{4,5,12}.

The apparent contradiction with other mapping studies^{13–15} probably results from the low resolution (~34 Å) in those experiments, combined with the use of long linkers and the size of the insertion proteins, which allow a difference density of >80 Å away from the insertion site. In this study, the docking was performed using a Laplacian filter^{11,16}, which was strictly required for the docking of individual parts of the structure. We suggest that this filtering approach be considered in future docking experiments of RyR domains and that a value for the docking contrast be reported. A complete validation of the docking and its relationship to previous attempts to locate the N-terminal region are available in the Supplementary Discussion. The docking shows that at least part of the intersubunit boundaries differs from those proposed previously¹⁷.

We mapped the positions of 33 disease-associated mutations in RyR1 and of 23 such mutations in RyR2 onto the RyR1 ABC structure (Fig. 3 and Supplementary Fig. 3). The positions can be grouped into three categories: those that are completely buried within a domain and thus probably cause local misfolding of the domain (6 mutations); those that lie at the interfaces between domains A, B and C within a subunit and thus probably destabilize the domain interactions (15 mutations); and those at interfaces with other RyR domains (35 mutations), including interfaces across subunits (Supplementary Table 2). Despite the presence of a large solvent-exposed surface area of the three domains within full-length RyR1, all mutations are either buried or at domain–domain interfaces. No mutations are simply exposed to solvent, where they would be less likely to cause a functional change, which further underlines the validity of the docking results (see also Supplementary Discussion).

Several mutations are concentrated at the boundary between domains A and C. This interface consists of a series of hydrogen-bonding and

¹Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada.

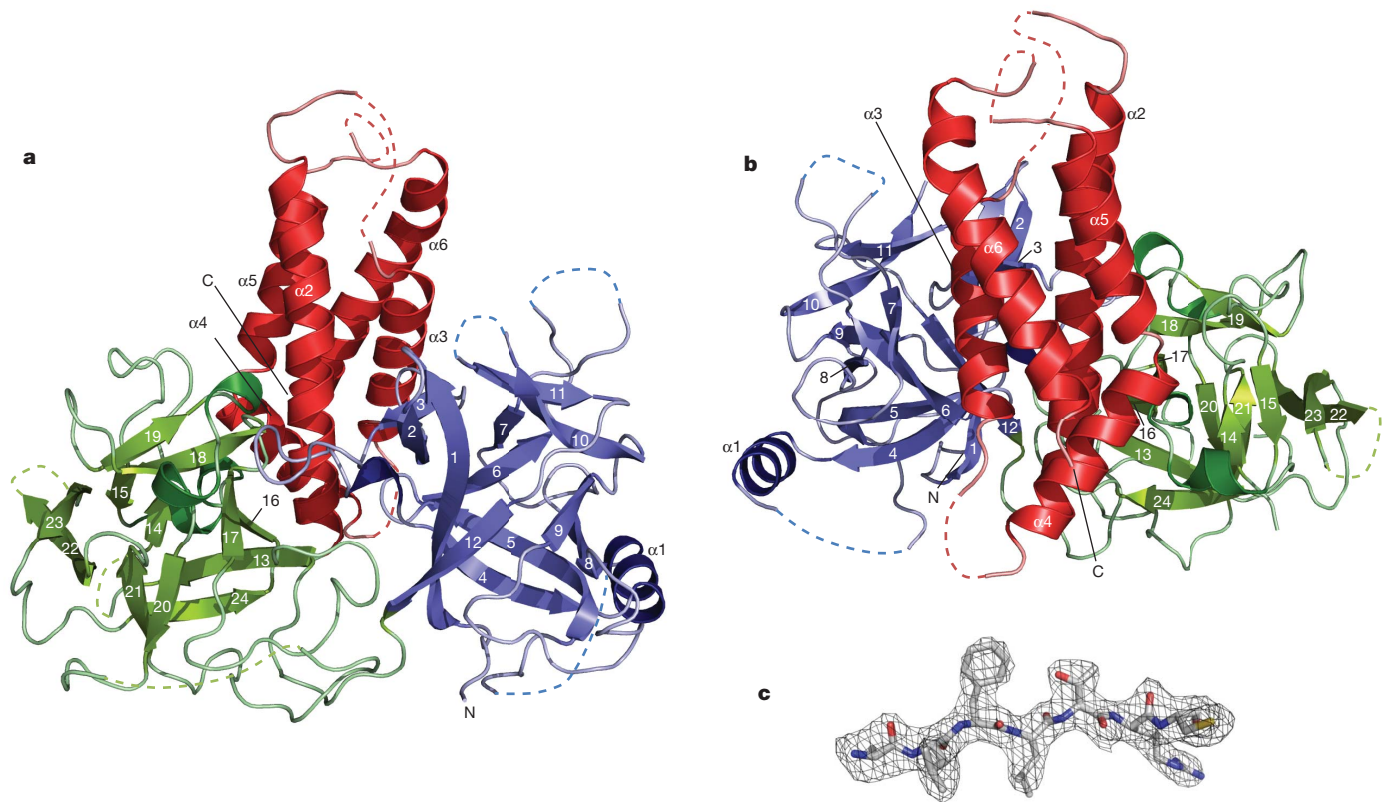


Figure 1 | Overall structure of the RyR1 A, B and C domains. **a**, Overall structure of rabbit RyR1 for amino acid residues 1–559. This portion of RyR1 consists of domain A (blue; 1–205), domain B (green; 206–394) and domain C (red; 395–532). The remaining 27 residues are not visible in the electron

charge-coupling interactions, including the amino acid R45 in domain A with D447 in domain C, and D61 and E40 in domain A with R402 in domain C (Fig. 3b). Both ionic pairs are the target of RyR1 mutations, through substitutions of R45, D61 and R402 (the residue numbering for rabbit RyR1 is used throughout for all disease-associated mutations) (Supplementary Table 2). Gel filtration and thermal melt experiments on the disease-associated RyR1 ABC mutant R402G (in which arginine has been substituted with glycine at position 402) show that this change does not cause overall misfolding but decreases the overall melting

density. Flexible loops are shown as dashed lines. For reference, α -helices ($\alpha 1$ – $\alpha 6$) and β -strands (1–24) are numbered. **b**, Different view of **a**. **c**, Typical omit difference density map in a randomly selected region, contoured at 3σ .

temperature by $>5^\circ\text{C}$ (Supplementary Fig. 4). It is therefore highly likely that this substitution simply destabilizes the A–C interface.

We have identified six individual interfaces between other parts of RyR1 and the A, B and C domains (Fig. 3c and Supplementary Fig. 3). One of these involves contacts with domains A and B across subunits (denoted interface 1). Depending on the cryo-EM map used, the domains are separated by a gap of ~ 5 – 11 Å, which can easily be bridged by two flexible loops in domain B (Fig. 3c). This variability in gap size may be due to differences in conformational purity of the

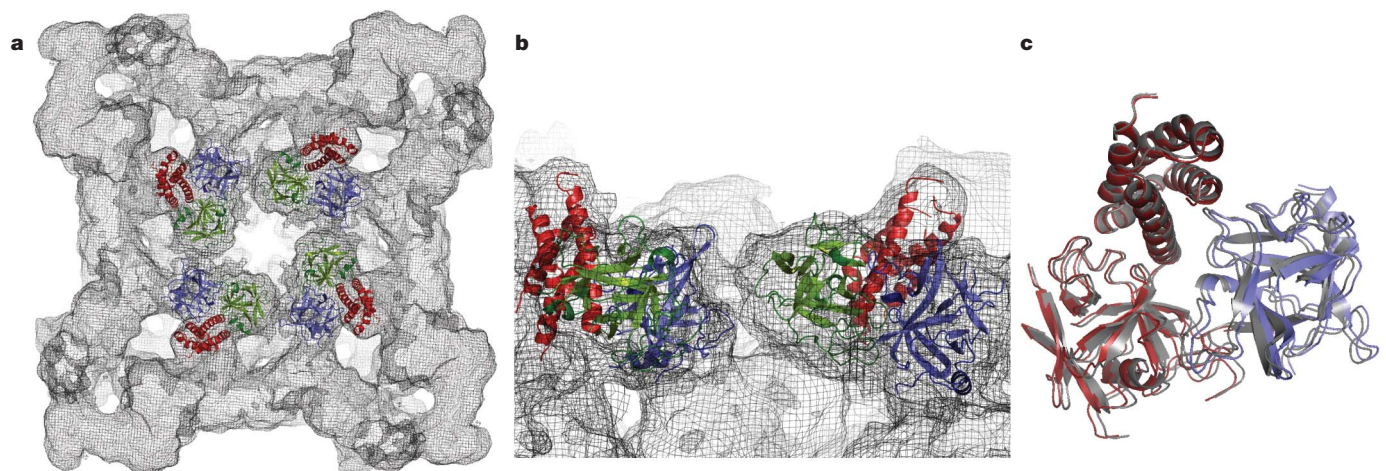


Figure 2 | Docking of RyR1 ABC in the 9.6 Å RyR1 cryo-EM map. **a**, View from the cytoplasmic side towards the endoplasmic reticulum, depicting domain A (blue), domain B (green) and domain C (red). **b**, Close-up lateral view from the four-fold symmetry axis. For clarity, only two monomers are

shown. **c**, Docking of domain A alone (blue) or domain BC alone (red) yields a near-perfect superposition on the docking position for ABC (grey). The probability of recapitulating the A–BC interface by chance is $<10^{-10}$ (see also Supplementary Discussion).

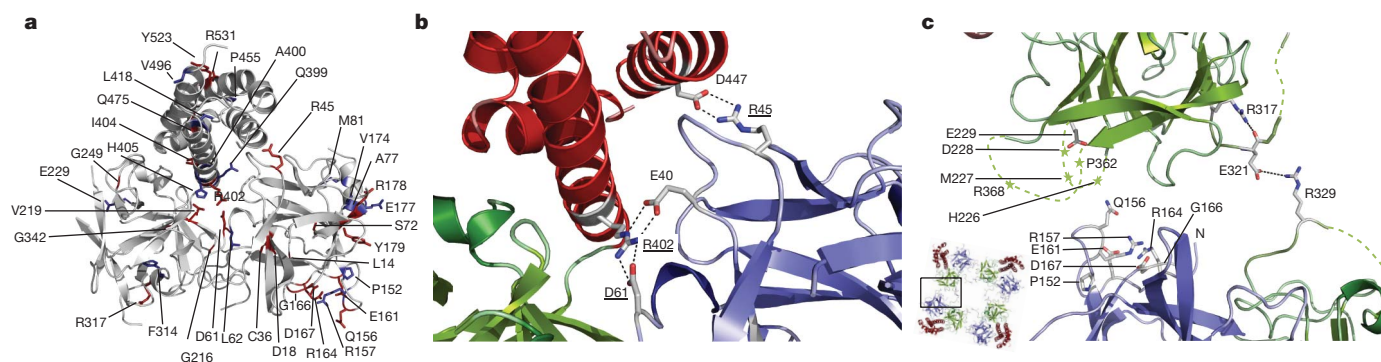


Figure 3 | Disease-associated mutations in RyR1 and RyR2. **a**, Position of disease-associated mutations in RyR1 (red) and additional such mutations in RyR2 (blue); regions not associated with disease (grey). For clarity, mutations in flexible loops are not shown. The view is similar to Fig. 1a. **b**, Salt bridges at the A–C interface. Domain A (blue), domain B (green) and domain C (red) are shown. Hydrogen bonds are represented by dashed lines. Underlined residues

RyR1 samples in the different studies⁹. Although most of the mutations on one side of this interface (domain B) are located in flexible loops, the docking shows the pairing of two oppositely charged residues (E321 and R329) on two adjacent B domains. The side chain of R329 is flexible in the crystal structure, but addition of the most common side-chain rotamer shows that the two residues can easily approach one another as close as 3.5 Å in the 10.3 Å map (Fig. 3c). Both residues are the target of disease-associated mutations, either directly (R329W in RyR1) or indirectly through mutation of a residue affecting the conformation of the main chain of E321 (R317L in RyR1, and R317W in RyR2) (Fig. 3c). Overall, 19 disease-associated mutations (in 16 positions) in RyR1 and RyR2 target this intersubunit interface, suggesting an important functional role for this contact area.

Other interfaces include areas lateral to domain A (interface 2) and domain C (interface 3) and areas 'below' domain A (interface 4), domain B (interface 5) and domain C (interface 6) (Supplementary Fig. 3). A large proportion of disease-associated mutations map to interface 4, connecting domain A to electron-dense columns extending towards the transmembrane domains (Supplementary Fig. 3a). This interface includes the single α -helix and the β 8– β 9 loop of domain A, two structural features that previously have been suggested to form important interactions with other RyR domains^{7,8}. The α -helix, which is the target of two disease-associated mutations and

are targets for disease-associated mutations. **c**, Interface 1 across ABC subunits, according to docking in the 10.3 Å RyR1 cryo-EM. Disease-associated mutation positions and E321 are labelled. Flexible loops are shown as coloured dashed lines. The R329 side chain is shown in its most common rotamer. R317 undergoes hydrogen bonding with the E321 main chain. The inset shows the relative view.

a deletion in RyR2 (refs 18,19), is also in contact with interface 2. Supplementary Table 2 gives a complete overview of the disease-associated mutations and their locations.

RyRs have been shown to be sensitive to redox modulation, a feature accomplished by several reactive cysteine residues that can be oxidized, nitrosylated or glutathionylated^{20–23}. In the context of full-length RyR1, C36 (rabbit RyR1 numbering) has been shown to be a target for glutathionylation²⁴. In domain A in isolation, C36 is highly flexible and accessible to solvent⁸; however, in the ABC crystal structure, it is completely buried by the β 16– β 17 loop of domain B (Fig. 4a and Supplementary Fig. 5). C36 is also the target for a disease-associated mutation (C36R) in RyR1. The A–B interface cannot accommodate an arginine residue or a glutathione moiety, both of which are bulkier than the cysteine residue. The fact that modification is possible in intact channels is strongly indicative that contacts between domains A and B are not static. In agreement with this idea, isothermal titration calorimetry experiments between domains A and BC that have been purified separately failed to show any interaction, suggesting that the contacts are inherently weak and can easily be broken (Supplementary Fig. 6).

Because docking of two individual parts of the crystal structure in the closed-state cryo-EM map recapitulates the same interface observed in the RyR1 ABC crystal structure (Fig. 2c), the ABC structure probably represents the relative domain orientation in the closed state of the

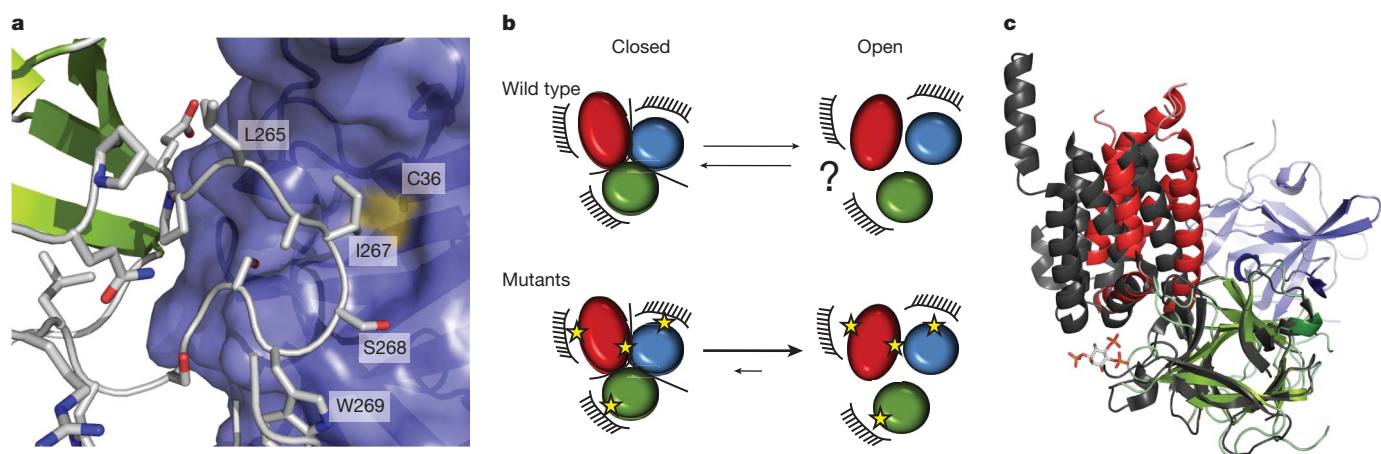


Figure 4 | The A–BC interface is labile. **a**, Surface of RyR1 domain A (blue) showing the C36 sulphhydryl (yellow), and representation of domain B (green). Selected residues from domain B are depicted as a grey stick structure, with oxygen atoms (red) and nitrogen atoms (blue) highlighted. C36 is inaccessible in the presence of domain B (Supplementary Fig. 5). **b**, Model showing the impact of disease-associated mutations (asterisks) on the motion of domains A, B and C (ovals). Interfaces with other RyR domains are indicated by black combs. The exact

nature and range of motion of these domains on opening of the channel are not known. Disease-associated mutations target multiple interfaces, thus facilitating motion of the domains. **c**, Superposition of the RyR1 ABC structure (coloured) with the InsP₃R BC structure (grey) bound to InsP₃ (stick representation; PDB ID: 1N4K), based on domain B of each structure. Treating the InsP₃R BC domain as a rigid unit shows a different relative orientation of domains B and C. This conformational difference might be induced by InsP₃ binding to InsP₃R.

channel. Although the existing cryo-EM maps of RyR1 in the open state are not publicly available for docking studies, they clearly show movements in the region where RyR1 ABC docks, including lateral movements away from the four-fold symmetry axis⁹. On the basis of the distribution of disease-associated mutations, we propose that the following occurs during the opening of RyR (Fig. 4b). Gating of the channel is allosterically coupled to movement of the A, B and C domains, either relative to one another or relative to other RyR domains. The distribution of disease-associated mutations at all observed interfaces suggests that all interfaces are prone to movement, including intersubunit contacts between domains A and B. Destabilization of any interface by disease-associated mutations facilitates the relative motion, resulting in a destabilized closed state and hence an increased probability of being open. This is compatible with previous functional studies showing that individual mutations either render the channels leaky or increase their sensitivity to activators⁶. The previously proposed 'zipper hypothesis' suggests that there is relative movement of the N-terminal and central disease hotspots²⁵. It is now clear, however, that the mutations do not target a single interface but instead affect multiple small zippers, both within and across the hotspots, as well as between the same hotspots on neighbouring subunits.

The N-terminal domains of another type of Ca^{2+} -release channel, inositol-1,4,5-trisphosphate receptors (InsP_3Rs)^{26,27}, show clear structural homology with the corresponding RyR1 domains (Fig. 4c and Supplementary Fig. 7). In InsP_3Rs , domain A is known to decrease the affinity of domain BC for InsP_3 (ref. 27), but the mechanism for this is not fully understood because an intact InsP_3R ABC structure is not available. Previous NMR and small-angle X-ray scattering data for $\text{InsP}_3\text{R1}$ ABC have suggested that the binding of InsP_3 changes the relative orientation of domains B and C, resulting in a destabilized A–BC interface²⁸. In this respect, it is notable that the relative position of domains B and C differs between the RyR1 ABC and InsP_3 -bound InsP_3R BC structures (Fig. 4c). Further experiments are needed to identify the exact domain movements that occur in both of these types of Ca^{2+} -release channel.

METHODS SUMMARY

Rabbit RyR1 (residues 1–559) was expressed at 18 °C in *Escherichia coli* Rosetta (DE3) pLacI cells as a fusion protein containing a hexahistidine tag, maltose-binding protein and a cleavage site for tobacco etch virus protease. The protein was purified to homogeneity using standard fast protein liquid chromatography (FPLC) techniques, and the tag was removed before crystallization trials. The protein was crystallized in 1.2–1.5 M $(\text{NH}_4)_2\text{SO}_4$, 0–15% glycerol (vol/vol) and 0.1 M Tris–Cl buffer, pH 7–9. Crystals were flash frozen in the same condition supplemented with 30% glycerol (vol/vol), and used for diffraction experiments at the Canadian Light Source beamline 08ID-1 and the Stanford Synchrotron Radiation Lightsource beamline 9-2. Data at 2.5 Å were processed with the program XDS²⁹, and R32 was indicated as the space group. The phase problem was solved using the N-terminal domain of RyR1 (refs 7,8) as a search model for the program Phaser³⁰. Supplementary Table 3 shows the data collection and refinement statistics. Docking experiments were performed by exhaustive six-dimensional searches using the programs Situs¹¹ and ADP_EM¹⁰. ThermoFluor experiments, used to determine the melting temperatures of various constructs, were performed as described previously⁸.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 April; accepted 3 September 2010.

Published online 3 November 2010.

- Giannini, G., Conti, A., Mammarella, S., Scrobogna, M. & Sorrentino, V. The ryanodine receptor/calcium channel genes are widely and differentially expressed in murine brain and peripheral tissues. *J. Cell Biol.* **128**, 893–904 (1995).
- Betzenhauser, M. J. & Marks, A. R. Ryanodine receptor channelopathies. *Pflügers Arch.* **460**, 467–480 (2010).
- Robinson, R., Carpenter, D., Shaw, M. A., Halsall, J. & Hopkins, P. Mutations in *RyR1* in malignant hyperthermia and central core disease. *Hum. Mutat.* **27**, 977–989 (2006).
- Ludtke, S. J., Serysheva, I. I., Hamilton, S. L. & Chiu, W. The pore structure of the closed RyR1 channel. *Structure* **13**, 1203–1211 (2005).

- Samsó, M., Wagenknecht, T. & Allen, P. D. Internal structure and visualization of transmembrane domains of the RyR1 calcium release channel by cryo-EM. *Nature Struct. Mol. Biol.* **12**, 539–544 (2005).
- MacLennan, D. H. & Chen, S. R. Store overload-induced Ca^{2+} release as a triggering mechanism for CPVT and MH episodes caused by mutations in *RyR* and *CASQ* genes. *J. Physiol. (Lond.)* **587**, 3113–3115 (2009).
- Amador, F. J. et al. Crystal structure of type I ryanodine receptor amino-terminal β -trefoil domain reveals a disease-associated mutation 'hot spot' loop. *Proc. Natl Acad. Sci. USA* **106**, 11040–11044 (2009).
- Lobo, P. A. & Van Petegem, F. Crystal structures of the N-terminal domains of cardiac and skeletal muscle ryanodine receptors: insights into disease mutations. *Structure* **17**, 1505–1514 (2009).
- Samsó, M., Feng, W., Pessah, I. N. & Allen, P. D. Coordinated movement of cytoplasmic and transmembrane domains of RyR1 upon gating. *PLoS Biol.* **7**, e85 (2009).
- Garzon, J. I., Kovacs, J., Abagyan, R. & Chacon, P. ADP-EM: fast exhaustive multi-resolution docking for high-throughput coverage. *Bioinformatics* **23**, 427–433 (2007).
- Chacón, P. & Wriggers, W. Multi-resolution contour-based fitting of macromolecular structures. *J. Mol. Biol.* **317**, 375–384 (2002).
- Serysheva, I. I., Hamilton, S. L., Chiu, W. & Ludtke, S. J. Structure of Ca^{2+} release channel at 14 Å resolution. *J. Mol. Biol.* **345**, 427–431 (2005).
- Liu, Z. et al. Three-dimensional reconstruction of the recombinant type 3 ryanodine receptor and localization of its amino terminus. *Proc. Natl Acad. Sci. USA* **98**, 6104–6109 (2001).
- Wang, R. et al. Localization of an NH_2 -terminal disease-causing mutation hot spot to the 'clamp' region in the three-dimensional structure of the cardiac ryanodine receptor. *J. Biol. Chem.* **282**, 17785–17793 (2007).
- Baker, M. L. et al. The skeletal muscle Ca^{2+} release channel has an oxidoreductase-like domain. *Proc. Natl Acad. Sci. USA* **99**, 12155–12160 (2002).
- Wriggers, W. & Chacón, P. Modeling tricks and fitting techniques for multiresolution structures. *Structure* **9**, 779–788 (2001).
- Serysheva, I. I. et al. Subnanometer-resolution electron cryomicroscopy-based domain models for the cytoplasmic region of skeletal muscle RyR channel. *Proc. Natl Acad. Sci. USA* **105**, 9610–9615 (2008).
- Bhuiyan, Z. A. et al. Expanding spectrum of human *RyR2*-related disease: new electrocardiographic, structural, and genetic features. *Circulation* **116**, 1569–1576 (2007).
- Marjamaa, A. et al. Search for cardiac calcium cycling gene mutations in familial ventricular arrhythmias resembling catecholaminergic polymorphic ventricular tachycardia. *BMC Med. Genet.* **10**, 12 (2009).
- Bellinger, A. M. et al. Hypernitrosylated ryanodine receptor calcium release channels are leaky in dystrophic muscle. *Nature Med.* **15**, 325–330 (2009).
- Durham, W. J. et al. RyR1 S-nitrosylation underlies environmental heat stroke and sudden death in Y522S RyR1 knockin mice. *Cell* **133**, 53–65 (2008).
- Eu, J. P., Sun, J., Xu, L., Stamler, J. S. & Meissner, G. The skeletal muscle calcium release channel: coupled O_2 sensor and NO signaling functions. *Cell* **102**, 499–509 (2000).
- Xu, L., Eu, J. P., Meissner, G. & Stamler, J. S. Activation of the cardiac calcium release channel (ryanodine receptor) by poly-S-nitrosylation. *Science* **279**, 234–237 (1998).
- Aracena-Parks, P. et al. Identification of cysteines involved in S-nitrosylation, S-glutathionylation, and oxidation to disulfides in ryanodine receptor type 1. *J. Biol. Chem.* **281**, 40354–40368 (2006).
- Ikemoto, N. & Yamamoto, T. Regulation of calcium release by interdomain interaction within ryanodine receptors. *Front. Biosci.* **7**, d671–d683 (2002).
- Bosanac, I. et al. Structure of the inositol 1,4,5-trisphosphate receptor binding core in complex with its ligand. *Nature* **420**, 696–700 (2002).
- Bosanac, I. et al. Crystal structure of the ligand binding suppressor domain of type 1 inositol 1,4,5-trisphosphate receptor. *Mol. Cell* **17**, 193–203 (2005).
- Chan, J. et al. Ligand-induced conformational changes via flexible linkers in the amino-terminal region of the inositol 1,4,5-trisphosphate receptor. *J. Mol. Biol.* **373**, 1269–1280 (2007).
- Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
- McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the staff at beamline 08ID-1 of the Canadian Light source, the Stanford Synchrotron Radiation Lightsource, K. Beam for the rabbit RyR1 clone, K. Lau for assistance with preparing the figures and the movie file, and C. Ahern, E. Moore and R. Pancaroglu for comments on the manuscript. F.V.P. is funded by the Canadian Institutes of Health Research (CIHR) and the Heart and Stroke Foundation of Canada, and is a CIHR new investigator and a Michael Smith Foundation for Health Research Scholar.

Author Contributions C.-C.T. expressed, purified and crystallized the protein, and collected diffraction data. P.A.L. cloned several initial constructs and assisted with the melting curve analysis. L.K. prepared the disease-associated mutation, purified the corresponding protein and measured the melting curves. F.V.P. designed and supervised the experiments, collected diffraction data, solved the structure, performed the docking experiments, and wrote the manuscript.

Author Information Atomic coordinates and structure factors for the RyR1 ABC structure have been deposited with the Protein Data Bank (<http://www.rcsb.org>) under accession code 2XOA. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.V.P. (filip.vanpetegem@gmail.com).

METHODS

Cloning, expression and purification. Rabbit RyR1 (amino acid residues 1–559) was cloned into a modified pET28 vector containing, in tandem, a hexahistidine tag, maltose-binding protein and a cleavage site for tobacco etch virus (TEV) protease followed by the construct of interest³¹. The R402G mutation was produced by using the QuikChange protocol (Stratagene). The protein was expressed at 18 °C in *Escherichia coli* Rosetta (DE3) pLacI cells (Novagen), induced at an absorbance of ~0.6 at 600 nm by adding 0.2 mM isopropyl- β -D-thiogalactoside (IPTG) and grown for 20 more hours before collection. Cells were lysed by sonication in buffer A (250 mM KCl and 10 mM HEPES, pH 7.4) with 25 μ g ml⁻¹ DNase I, 25 μ g ml⁻¹ lysozyme, 14 mM β -mercaptoethanol and 1 mM phenylmethylsulphonyl fluoride. The lysate was applied to a 25 ml Poros MC column (Tosoh Bioscience), washed with five column volumes of buffer A and five column volumes of buffer A plus 2% (vol/vol) buffer B (250 mM KCl and 500 mM imidazole, pH 7.4) and eluted with 30% (vol/vol) buffer B. The protein was dialysed against buffer A plus 14 mM β -mercaptoethanol and then applied to a 25 ml amylose column (New England Biolabs). The protein was then washed with ten column volumes of buffer A plus 14 mM β -mercaptoethanol and eluted with buffer C (buffer A plus 10 mM maltose and 14 mM β -mercaptoethanol). The protein was dialysed overnight against buffer D (50 mM KCl, 10 mM HEPES, pH 7.4, and 14 mM β -mercaptoethanol) at 4 °C and was cleaved simultaneously with recombinant TEV protease. The sample was then applied to another 25 ml Poros MC column in buffer D, and the flow through was collected and applied to a 6 ml Resource Q column (GE Healthcare). The protein was eluted with a gradient of 0% to 40% buffer E (1.5 M KCl, 10 mM HEPES, pH 7.4, and 14 mM β -mercaptoethanol). The sample was run on a HiLoad 16/10 Phenyl Sepharose HP column (GE Healthcare) in buffer E and eluted with a gradient of 0% to 50% buffer D. Last, the sample was run on a HiLoad 16/60 Superdex 200 prep grade (GE Healthcare) gel filtration column using buffer A plus 14 mM β -mercaptoethanol. The protein sample was exchanged into 50 mM KCl, 10 mM HEPES, pH 7.4, and 5 mM dithiothreitol and was then concentrated to 10–20 mg ml⁻¹ using Amicon concentrators (30 kDa molecular mass cutoff; Millipore). The purified protein was stored at -80 °C.

Crystallization, data collection and structure solution. Crystals were obtained by hanging-drop vapour diffusion. Crystals appeared in a solution of 1.2–1.5 M (NH₄)₂SO₄, 0–15% glycerol (vol/vol) and 0.1 M Tris-Cl, pH 7–9. Crystals were transferred to the same condition supplemented with 30% glycerol (vol/vol), and flash frozen in liquid nitrogen before data collection. Data sets were collected at the Canadian Light Source (CLS) beamline 08ID-1 and the Stanford Synchrotron Radiation Lightsource beamline 9-2. The data set for final refinement was collected at the CLS at a wavelength of 0.9795 Å and a temperature of 100 K, and was processed using the XDS program²⁹. The structure of RyR1 domain A^{7,8} (PDB

ID: 3ILA) was used as a search model in molecular replacement with the Phaser program³⁰. The resultant electron-density maps were used for the initial placement of four α -helices and were then used for automated structure building implemented in the program ARP/wARP version 7.1 (ref. 32). The model was further completed and refined using alternating cycles of manual building in the programs Coot³³ and REFMAC version 5 (ref. 34). The geometry of the model was validated using the PROCHECK program³⁵, showing no residues in disallowed regions of the Ramachandran plots and with 87.5% in the core region. Residues with missing densities were not modelled. Supplementary Table 3 shows the data collection and refinement statistics.

Docking into cryo-EM maps. The resultant structure was docked into three RyR1 cryo-EM maps (Electron Microscopy Data Bank ID: 1275, 5014 and 1274), solved at a reported 9.6 Å (ref. 4), 10.3 Å (ref. 5) and 14 Å (ref. 12), respectively. We used extensive six-dimensional searches as implemented in the programs Situs¹¹ and ADP_EM¹⁰. Laplacian filtering was used to prevent a multitude of spurious solutions due to migration to regions with high electron density¹⁶. The search model was blurred to the same resolution as the cryo-EM map. No additional filtering was performed on the map. An angular step of 3° was used for the initial six-dimensional search, followed by Powell minimization. A confidence interval and docking precision estimate were obtained using methods described previously^{36,37}, making use of Fisher's *z* transform of the correlation coefficient and an estimated $\sigma(z)$ from the map volume and resolution.

Thermal melt analysis. Thermal melting curves for purified proteins were obtained from ThermoFluor experiments as described previously⁸. The midpoint of the transitions was used as an indicator of overall protein stability.

31. Van Petegem, F., Clark, K. A., Chatelain, F. C. & Minor, D. L. Jr. Structure of a complex between a voltage-gated calcium channel β -subunit and an α -subunit domain. *Nature* **429**, 671–675 (2004).
32. Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature Protocols* **3**, 1171–1179 (2008).
33. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
34. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
35. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
36. Volkman, N. Confidence intervals for fitting of atomic models into low-resolution densities. *Acta Crystallogr. D* **65**, 679–689 (2009).
37. Volkman, N. & Hanein, D. Docking of atomic models into reconstructions from electron microscopy. *Methods Enzymol.* **374**, 204–225 (2003).

A redox switch in angiotensinogen modulates angiotensin release

Aiwu Zhou¹, Robin W. Carrell², Michael P. Murphy³, Zhenquan Wei¹, Yahui Yan¹, Peter L. D. Stanley¹, Penelope E. Stein², Fiona Broughton Pipkin⁴ & Randy J. Read¹

Blood pressure is critically controlled by angiotensins¹, which are vasopressor peptides specifically released by the enzyme renin from the tail of angiotensinogen—a non-inhibitory member of the serpin family of protease inhibitors^{2,3}. Although angiotensinogen has long been regarded as a passive substrate, the crystal structures solved here to 2.1 Å resolution show that the angiotensin cleavage site is inaccessibly buried in its amino-terminal tail. The conformational rearrangement that makes this site accessible for proteolysis is revealed in our 4.4 Å structure of the complex of human angiotensinogen with renin. The co-ordinated changes involved are seen to be critically linked by a conserved but labile disulphide bridge. Here we show that the reduced unbridged form of angiotensinogen is present in the circulation in a near 40:60 ratio with the oxidized sulphydryl-bridged form, which preferentially interacts with receptor-bound renin. We propose that this redox-responsive transition of angiotensinogen to a form that will more effectively release angiotensin at a cellular level contributes to the modulation of blood pressure. Specifically, we demonstrate the oxidative switch of angiotensinogen to its more active sulphydryl-bridged form in the maternal circulation in pre-eclampsia—the hypertensive crisis of pregnancy that threatens the health and survival of both mother and child.

The octapeptide angiotensin II that constricts arterial blood vessels is released from a precursor decapeptide, angiotensin I, by the angiotensin converting enzyme, ACE. Although studies of hypertension and its therapy have focused on ACE and its inhibitors, the initiating and rate-limiting step in the angiotensin pathway is the preceding release of angiotensin I, by the enzyme renin, from the N terminus of the much larger protein, angiotensinogen^{1,4} (Fig. 1a). Previous attempts to determine the structure of plasma angiotensinogen were frustrated by its heterogeneity, and in order to obtain diffraction-quality crystals, we produced recombinant non-glycosylated angiotensinogens from mouse, rat and human by expression in *Escherichia coli*. Subsequent crystallization and X-ray diffraction yielded data sets that were solved as detailed in Methods. We determined the mouse structures at higher resolution (2.3, 2.1, 2.95 Å), but the general features are preserved in the 3.15 Å rat and 3.3 Å human structures (as shown separately and with statistics in Supplementary Fig. 2 and Supplementary Table 1).

With the exception of the exposed and hence mobile reactive centre loop (residues 408–425), the overall results show the complete fold of the human molecule, including the full N-terminal extension with the renin cleavage site (Fig. 1a). Although angiotensinogen has only 22% sequence identity to its closest relatives amongst other serpins³, it substantially retains the typical serpin fold (Supplementary Fig. 2). Of special interest, the structure shows that Met 235, the common polymorphism of which has been linked to hypertension, has a freely extended side-chain and is situated on an external peptide loop (hF-s3A) distant from the main functional sites on the molecule. This strengthens previous deductions^{5,6} that the predisposition to hypertension results from the small increase in

concentration of the polymorphic angiotensinogen rather than a change in its function. The 63-residue N-terminal tail of angiotensinogen is seen as an ordered superstructure, anchored by two new helices, and with the renin-cleavage site, Leu 10–Val 11 in humans, held in an inaccessibly buried position. The change in fold allowing access to the buried cleavage site became apparent when we solved the structure of the complex of angiotensinogen and renin. Crystals of the initiating complex formed by human angiotensinogen with recombinantly inactivated human renin (Asp292Ala) gave X-ray diffraction to 4.4 Å resolution. The structure was solved by molecular replacement, as detailed in Methods. Although at low resolution, the structure of the complex clearly shows the mode of binding of renin and the associated localized changes in fold (Fig. 1b, Supplementary Fig. 3).

In addition to the predicted^{7,8} intimate interaction between the N-terminal substrate (angiotensin I) region of angiotensinogen and the active site cleft of renin, there is a substantial contact surface of 670 Å² between the bodies of the two proteins, primarily hydrophobic in nature. Helix A of angiotensinogen lies across the active site cleft of renin, and helix C and the CD-loop contact the N-terminal lobe of renin. The body of angiotensinogen substantially retains its fold except for a displacement of the CD-loop, which is enlarged and conserved in the angiotensinogens and would otherwise sterically block the binding of renin. The 10 Å shift of this CD loop is accompanied by a 10–20 Å extension and movement of the N-terminal substrate peptide of angiotensinogen into the active site cleft of renin with, as clearly seen in the Supplementary Video, the two concerted movements being linked by a disulphide bridge.

The functional significance of this Cys 18–Cys 138 bridge linking the angiotensin-containing tail of angiotensinogen to the body of the molecule had been indicated in earlier biochemical studies⁹. Human angiotensinogen contains four cysteines but those forming the Cys 18–Cys 138 linkage are the only two conserved in all species^{9,10}. We now show (Fig. 2a) that this disulphide linkage is labile, with a redox titration of isolated angiotensinogen indicating a reduction potential for this dithiol/disulphide equilibrium of –230 mV at 25 °C, pH 7 (Supplementary Fig. 4). That reduction does result in the breaking of the 18–138 bridge was confirmed in the crystal structure of reduced mouse angiotensinogen, with the likely structural and hence functional consequences being indicated by a change of thermal stability, from 62.6 °C for the constrained oxidized form to 57 °C for the relaxed reduced form (Supplementary Fig. 5). Angiotensinogen is predictably secreted in the oxidized bridged form after passing through the endoplasmic reticulum, but will then be reduced by interaction with selective thiol-reductant systems in the plasma. The presence of the reduced as well as oxidized forms of angiotensinogen in the plasma is shown in Fig. 2b, with a remarkably consistent reduced-to-oxidized proportion, near 40:60 independent of gender or age, as shown below. Although this is a stable ratio in the plasma pool, the redox poise of the transition will predictably allow the reversible switch from the oxidized to reduced form in focal tissues or vascular beds.

¹Department of Haematology, Cambridge Institute for Medical Research, University of Cambridge, Hills Road, Cambridge CB2 0XY, UK. ²Department of Medicine, Cambridge Institute for Medical Research, University of Cambridge, Hills Road, Cambridge CB2 0XY, UK. ³MRC Mitochondrial Biology Unit, Hills Road, Cambridge CB2 0XY, UK. ⁴Department of Obstetrics and Gynaecology, University of Nottingham, City Hospital, Nottingham NG5 1PB, UK.

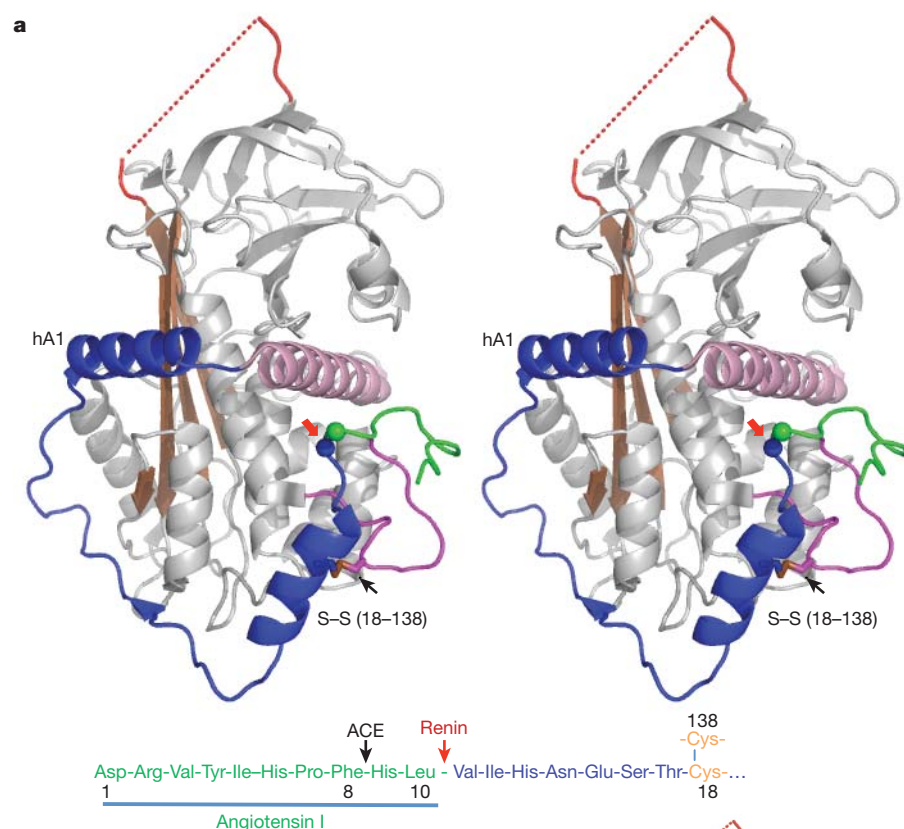


Figure 1 | Angiotensinogen and its complex with renin. **a**, Stereo image of human angiotensinogen: the serpin template is shown in grey and helix A in purple with the A-sheet in brown, the unresolved reactive loop dotted in red, and the CD loop containing Cys 138 in magenta. The amino-tail is in blue with the new helix A1 and a second helix A2 containing Cys 18 (linked in brown to Cys 138); the terminal angiotensin I segment is in green with the renin-cleavage site shown as green and blue balls. The sequence below (same colour coding) also indicates the subsequent cleavage by angiotensin converting enzyme (ACE) releasing the octapeptide angiotensin II. **b**, Angiotensin in the initiating complex with inactivated (Asp292Ala) renin (left), and on right superimposed on the unreacted form (brown, renin-cleavage site red arrow) showing the displacement of the CD loop and the movement of the N-terminal peptide (visible to Cys 18, unresolved residues 19–30 blue-dotted, renin-cleavage site blue arrow), into the active cleft of renin (see also Supplementary Fig. 3).

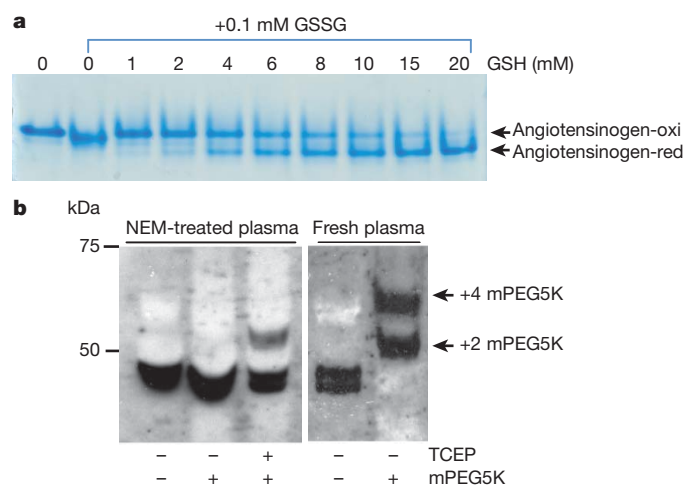
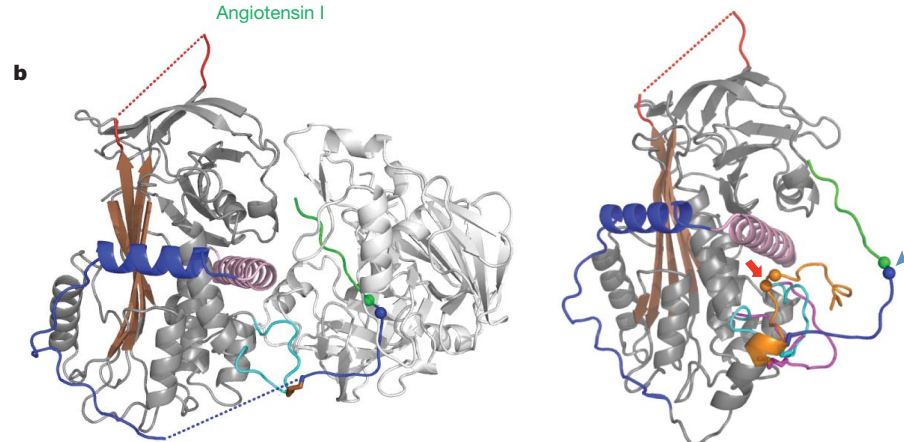


Figure 2 | Reduced angiotensinogen in plasma. **a**, Ready reduction of the 18–138 disulphide is seen in the native PAGE of oxidized recombinant human angiotensinogen (angiotensinogen-oxi) incubated at 23°C for 10 h with glutathione disulphide (GSSG, 0.1 mM) and glutathione (GSH, 0–20 mM) and subsequent blockage of the reduced form (angiotensinogen-red) with iodoacetate (faster moving band). **b**, Western blot SDS-PAGE of plasma exposed to N-ethylmaleimide (NEM) and its high molecular weight polyethylene-glycol adduct (mPEG5K). Blockage of the free SH groups at Cys 232 and 308 with NEM and subsequent reduction with tris(2-carboxyethyl) phosphine (TCEP) of the 18–138 bridge and incubation with mPEG5K (lane 3) reveals 2 SH. With fresh unreduced plasma, incubation with mPEG5K (lane 5) reveals the presence of the two angiotensinogen forms—oxidized with 2 available SH groups (+2 mPEG5K) and reduced with 4 SH (+4 mPEG5K), giving the consistent proportions of the two forms in human plasma seen in Fig. 4a.

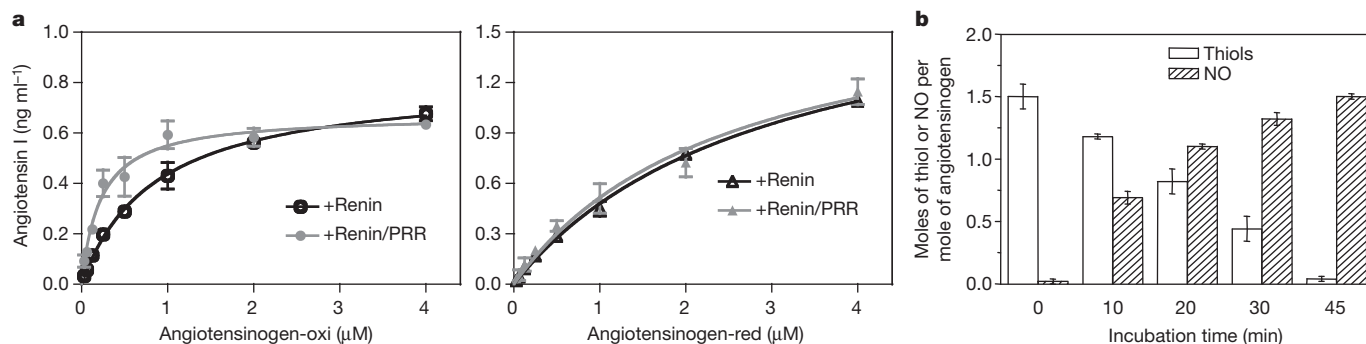


Figure 3 | Kinetics and nitrosylation. **a**, Reactions of human renin and fully glycosylated angiotensinogen (oxidized and reduced, ‘-oxi’ and ‘-red’) in the presence or absence of prorenin receptor (PRR) with angiotensin I concentration (after 200-fold dilution) measured by an EIA kit. Detailed kinetic parameters are given in Table 1. **b**, The time course of modification of the

18–138 thiols of reduced angiotensinogen on incubation at 37 °C with the S-nitrosothiol donor SNAP, showing the ready blocking and conversion of the two SH groups (Thiols) to S-nitrosothiols (NO). Values are averages of three independent measurements (mean \pm s.d., $n = 3$).

The critical interactions of renin and angiotensinogen, with respect to the control of blood pressure, are now believed to occur in the renal tubules and tissue¹¹. At these sites a third component, the cell-surface prorenin receptor, binds renin to give a considerable enhancement of its catalytic cleavage of angiotensinogen¹². To assess the functional consequences of the oxidative transition, we recombinantly expressed and isolated fully glycosylated human angiotensinogen, human renin and human prorenin receptor. Incubation of the reduced and oxidized forms of angiotensinogen with renin (Fig. 3a, Table 1) shows that the prorenin receptor although having little effect on the reduced form gives a fourfold increase in the renin-binding affinity (K_m) of the oxidized form, with a consequent fourfold increase in the catalytic release of angiotensin. This preferential binding and activation of the oxidized form of angiotensinogen by the prorenin receptor will particularly occur on cell surfaces, thus allowing a redox-sensitive and focal modulation of angiotensin release in specific tissues.

Although we demonstrate here the direct effects of the Cys 18–Cys 138 reduction on the substrate activity of angiotensinogen, the subsequent ligand modification of reduced disulphides is now also known to be widely involved in the modulation of activity in other proteins¹³. The potential for such modifications is pertinently¹⁴ illustrated with a known hypotensive agent, nitric oxide. As shown in Fig. 3b, incubation of reduced angiotensinogen with the S-nitrosothiol donor SNAP^{15,16} leads to the stoichiometric loss of the 18 and 138 thiols, confirming the potential nitrosylation and hence blockage of both cysteines. Taken together with the structural evidence, these findings of the consequences of the redox transition of angiotensinogen strongly indicate an as yet unrecognized modulatory mechanism at the commencement of the principal pathway controlling vasoconstriction.

Previous indications of a redox-sensitive contribution to the control of blood pressure came from observations of the association of oxidative stress with the onset of episodic hypertension^{17,18}. Our suspicion that this could result from the oxidative conversion of angiotensinogen to its more active bridged form was difficult to confirm, as the critical release of angiotensins is likely to occur focally in renal tissue rather than in the plasma pool. A more overt occurrence however is the oxidative stress arising from placental dysfunction^{14,19}. This underlies^{20,21} the hypertension that commonly complicates pregnancy and progresses in a quoted 2–7% of all pregnancies to the development of pre-eclampsia, a leading

cause of perinatal and maternal mortality and morbidity in the developed world²². To determine, as proof of principle, if oxidative changes occurred in angiotensinogen in pre-eclampsia, we examined samples from a previous study of pre-eclamptic women²³.

Twenty-four EDTA plasma samples, stored at –80 °C, 12 from women with strictly defined pre-eclampsia and 12 from gestation matched normotensive controls, were referred for analysis with a blinded coding. The proportion of reduced to oxidized angiotensinogen in each sample was then assessed by electrophoresis with western blotting and the gels compared with the control series in Fig. 4a. Visual examination of the 24 maternal samples (Fig. 4b) clearly indicated the presence of samples with a decreased proportion of reduced to oxidized angiotensinogen, allowing the ready recognition of 8 of the 12 pre-eclamptic plasmas before the breaking of the code. Unequivocal evidence of the occurrence of the oxidative transition in pre-eclampsia came from the samples with strikingly obvious decreases in reduced angiotensinogen, as asterisked in Fig. 4b, all of which were revealed to be from pre-eclamptics (see also Supplementary Fig. 6). Subsequent quantitative measurements (Fig. 4c) confirm the visual observations from the gel and also the findings in studies of other plasma markers^{23,24} of an overlapping range of oxidative changes in pregnancy plasma—from mild in normotensives to severe in pre-eclampsia. These changes in the plasma pool will be much more marked in the tissues. The more effective cleavage of the oxidized form of angiotensinogen by cell-bound renin in these tissues thus provides a contributory causative link between the oxidative changes in pregnancy and the onset of the hypertension that is a defining feature of pre-eclampsia.

In conclusion, a central message from our findings is that the release of angiotensin requires a change in conformation in angiotensinogen to allow the access, as well as complementary binding, of renin. At the focus of this change is a labile disulphide bridge that readily undergoes a transition to a more active oxidized form and demonstrably does so in pre-eclampsia. Evidence that the fourfold change in activity that occurs on oxidation is a sufficient cause of pre-eclamptic hypertension comes from a fortuitous experiment of nature. A mutation at the renin cleavage site in angiotensinogen, Leu10Phe, has been previously detected in five heterozygotes, each identified because of an association with hypertension and specifically with hypertension in pregnancy. Investigation of the index case of a 17-year-old woman with pre-eclampsia²⁵ showed that the mutation resulted in an increased catalytic efficiency of cleavage of the angiotensinogen by renin, comparable to the change observed here with the transition of angiotensinogen from the reduced to oxidized form. Thus there is now a clear chain of evidence linking the oxidative changes in the placenta with the hypertension in pregnancy that leads on to the development of pre-eclampsia. The findings here will reinforce efforts to develop anti-oxidant therapies for pre-eclampsia²⁶ and will also open a means of monitoring the progress and management of the disorder. In a broader context, we draw attention to the hitherto

Table 1 | Detailed kinetics

Glycosylated ANT	[Renin]	K_m (μM)	k_{cat} (s ^{–1})	k_{cat}/K_m (s ^{–1} μM ^{–1})
ANT-oxi	0.1 nM	0.86 \pm 0.05	0.57 \pm 0.02	0.66 (100%)
ANT-red	0.1 nM	2.9 \pm 0.3	1.3 \pm 0.1	0.45 (68%)
ANT-oxi	0.1 nM + PRR	0.22 \pm 0.09	0.47 \pm 0.1	2.1 (318%)
ANT-red	0.1 nM + PRR	2.5 \pm 0.5	1.3 \pm 0.2	0.52 (79%)

ANT, angiotensinogen; PRR, prorenin receptor.

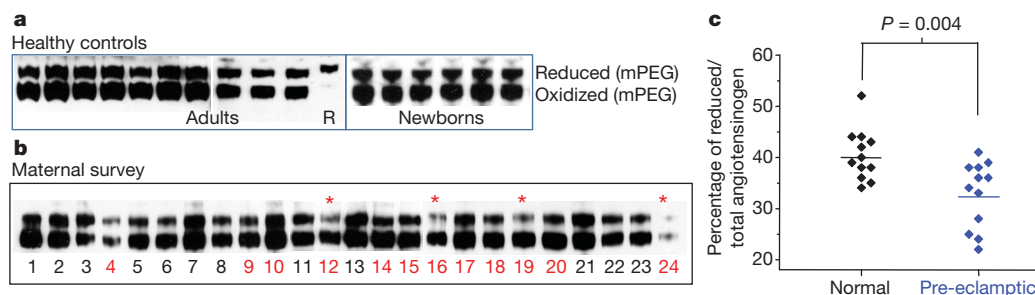


Figure 4 | Reduced and oxidized plasma angiotensinogen in pregnancy. Western blot SDS-PAGE of plasma exposed to N-ethylmaleimide-polyethylene-glycol adduct (mPEG). **a**, Plasma samples from adult controls show remarkably constant 40:60 proportions of the two forms, as do cord (venous umbilical) plasmas from healthy newborns—taken and stored with the maternal plasmas. R, reduced plasma treated with mPEG. **b**, Maternal plasma samples from pre-eclampsia and matched normotensive controls, as

sequentially referred with blinded coding. The recognizably lower proportion of reduced angiotensinogen in pre-eclamptic plasmas was confirmed on breaking of the coding (normotensives black numerals, pre-eclampsics red, asterisked four samples self-evidently below the mean in **c**). **c**, Percentages of the reduced-form in pre-eclampsics (blue), mean $32 \pm 6\%$; and matched normotensive controls (black), mean $40 \pm 5\%$ (means \pm s.d., $n = 12$, $P = 0.004$ by Student's t -test). See also Supplementary Fig. 6.

neglected contribution of angiotensinogen to the multifactorial processes that modulate blood pressure in humans.

METHODS SUMMARY

Crystals of angiotensinogen and the complex of human angiotensinogen with deglycosylated renin-Asp292Ala were grown at room temperature under conditions listed in Supplementary Table 1. Data sets were collected at 100 K at the Daresbury and Diamond synchrotron light sources and processed with the CCP4 program suite. The structures of human, rat and mouse angiotensinogen were solved by molecular replacement in Phaser²⁷, using an ensemble of distantly related serpins as the search model with additional phase information obtained from a GdCl₃ derivative of human angiotensinogen. The structure of the complex with renin was solved by molecular replacement using renin from PDB entry 2BKS²⁸ and its peptide complex from PDB entry 1SMR⁷. The structures were completed by refinement in Refmac5²⁹ and rebuilding in Coot³⁰. Further details related to crystal structures and refinement, recombinant expression, thiol titration and nitrosylation and the assessment of sulphhydryl blockage and oxidation, the determination of redox potentials, renin kinetics and the demonstration of reduced and oxidized forms of angiotensinogen, are given in Supplementary Methods.

Received 8 December 2009; accepted 4 August 2010.

Published online 6 October 2010.

- Fyhrquist, F. & Saijonmaa, O. Renin-angiotensin system revisited. *J. Intern. Med.* **264**, 224–236 (2008).
- Stein, P. E., Tewkesbury, D. A. & Carrell, R. W. Ovalbumin and angiotensinogen lack serpin S-R conformational change. *Biochem. J.* **262**, 103–107 (1989).
- Huber, R. & Carrell, R. W. Implications of the three-dimensional structure of α_1 -antitrypsin for structure and function of serpins. *Biochemistry* **28**, 8951–8966 (1989).
- Dickson, M. E. & Sigmund, C. D. Genetic basis of hypertension: revisiting angiotensinogen. *Hypertension* **48**, 14–20 (2006).
- Kim, H. S. *et al.* Genetic control of blood pressure and the angiotensinogen locus. *Proc. Natl Acad. Sci. USA* **92**, 2735–2739 (1995).
- Inoue, I. *et al.* A nucleotide substitution in the promoter of human angiotensinogen is associated with essential hypertension and affects basal transcription *in vitro*. *J. Clin. Invest.* **99**, 1786–1797 (1997).
- Dealwis, C. G. *et al.* X-ray analysis at 2.0 Å resolution of mouse submaxillary renin complexed with a decapeptide inhibitor CH-66, based on the 4–16 fragment of rat angiotensinogen. *J. Mol. Biol.* **236**, 342–360 (1994).
- Nasir, U. M. *et al.* Two peaks in pH dependence of renin-angiotensinogen reaction. *Biosci. Biotechnol. Biochem.* **62**, 338–340 (1998).
- Streetfield-James, R. M. *et al.* Angiotensinogen cleavage by renin: importance of a structurally constrained N-terminus. *FEBS Lett.* **436**, 267–270 (1998).
- Gimenez-Roqueplo, A. P., Celerier, J., Schmid, G., Corvol, P. & Jeunemaitre, X. Role of cysteine residues in human angiotensinogen. Cys232 is required for angiotensinogen-pro major basic protein complex formation. *J. Biol. Chem.* **273**, 34480–34487 (1998).
- Kobori, H., Nangaku, M., Navar, L. G. & Nishiyama, A. The intrarenal renin-angiotensin system: from physiology to the pathobiology of hypertension and kidney disease. *Pharmacol. Rev.* **59**, 251–287 (2007).
- Nguyen, G. *et al.* Pivotal role of the renin/prorenin receptor in angiotensin II production and cellular responses to renin. *J. Clin. Invest.* **109**, 1417–1427 (2002).
- Brandes, N., Schmitt, S. & Jakob, U. Thiol-based redox switches in eukaryotic proteins. *Antioxid. Redox Signal.* **11**, 997–1014 (2009).
- Myatt, L. Review: reactive oxygen and nitrogen species and functional adaptation of the placenta. *Placenta* **31** (suppl. A), S66–S69 (2010).
- Dahm, C. C., Moore, K. & Murphy, M. P. Persistent S-nitrosation of complex I and other mitochondrial membrane proteins by S-nitrosothiols but not nitric oxide or

- peroxynitrite: implications for the interaction of nitric oxide with mitochondria. *J. Biol. Chem.* **281**, 10056–10065 (2006).
- Stamler, J. S., Lamas, S. & Fang, F. C. Nitrosylation: the prototypic redox-based signaling mechanism. *Cell* **106**, 675–683 (2001).
- Wilcox, C. S. Oxidative stress and nitric oxide deficiency in the kidney: a critical link to hypertension? *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **289**, R913–R935 (2005).
- Harrison, D. G. & Gongora, M. C. Oxidative stress and hypertension. *Med. Clin. North Am.* **93**, 621–635 (2009).
- Burton, G. J. & Jauniaux, E. Placental oxidative stress: from miscarriage to preeclampsia. *J. Soc. Gynecol. Invest.* **11**, 342–352 (2004).
- Hubel, C. A. Oxidative stress in the pathogenesis of preeclampsia. *Proc. Soc. Exp. Biol. Med.* **222**, 222–235 (1999).
- Sedeek, M. *et al.* Role of reactive oxygen species in hypertension produced by reduced uterine perfusion in pregnant rats. *Am. J. Hypertens.* **21**, 1152–1156 (2008).
- Sibai, B., Dekker, G. & Kupferminc, M. Pre-eclampsia. *Lancet* **365**, 785–799 (2005).
- Mistry, H. D., Wilson, V., Ramsay, M. M., Symonds, M. E. & Broughton Pipkin, F. Reduced selenium concentrations and glutathione peroxidase activity in preeclamptic pregnancies. *Hypertension* **52**, 881–888 (2008).
- Bulgan Kilicdag, E. *et al.* Oxidant-antioxidant system changes relative to placental-umbilical pathology in patients with preeclampsia. *Hypertens. Pregnancy* **24**, 147–157 (2005).
- Inoue, I. *et al.* A mutation of angiotensinogen in a patient with preeclampsia leads to altered kinetics of the renin-angiotensin system. *J. Biol. Chem.* **270**, 11430–11436 (1995).
- Hoffmann, D. S. *et al.* Chronic tempol prevents hypertension, proteinuria, and poor fetal-placental outcomes in BPH/5 mouse model of preeclampsia. *Hypertension* **51**, 1058–1065 (2008).
- McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Powell, N. A. *et al.* Benzyl ether structure-activity relationships in a series of ketopiperazine-based renin inhibitors. *Bioorg. Med. Chem. Lett.* **15**, 4713–4716 (2005).
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature. A summary figure is also included.

Acknowledgements This study was funded by the British Heart Foundation, the Wellcome Trust, the Isaac Newton Trust of the University of Cambridge and the UK Medical Research Council, with the support of the Daresbury SRS and Diamond Light Source. We acknowledge the assistance of G. Bunkóczi in data processing, H. Hong for the HPLC measurements of angiotensin, and T. Prime for analysing S-nitrosothiols.

Author Contributions This paper is a culmination of a 20-year study initiated by R.W.C. with P.E.S.; R.W.C. guided the work throughout and wrote the paper with input from A.Z. and R.J.R.; A.Z. designed, performed and analysed all the work reported here, with Z.W., Y.Y. and, for redox studies, with M.P.M.; P.E.S. with P.L.D.S. carried out preceding studies, and F.B.P. provided samples and advice for the pre-eclamptic studies. R.J.R. advised on all aspects of this study and solved the structures. All authors discussed the results and commented on the manuscript.

Author Information Ethical approval was obtained for the use of all patient and control plasmas. Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession codes 2WXX, 2WXY, 2WXZ, 2WXW, 2WYO, 2WY1 and 2XOB. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.W.C. (rw1000@cam.ac.uk) or A.Z. (awz20@cam.ac.uk).

Reducing excessive GABA-mediated tonic inhibition promotes functional recovery after stroke

Andrew N. Clarkson^{1*†}, Ben S. Huang^{1,2*}, Sarah E. MacIsaac¹, Istvan Mody^{1,2,3} & S. Thomas Carmichael¹

Stroke is a leading cause of disability, but no pharmacological therapy is currently available for promoting recovery. The brain region adjacent to stroke damage—the peri-infarct zone—is critical for rehabilitation, as it shows heightened neuroplasticity, allowing sensorimotor functions to re-map from damaged areas^{1–3}. Thus, understanding the neuronal properties constraining this plasticity is important for the development of new treatments. Here we show that after a stroke in mice, tonic neuronal inhibition is increased in the peri-infarct zone. This increased tonic inhibition is mediated by extrasynaptic GABA_A receptors and is caused by an impairment in GABA (γ-aminobutyric acid) transporter (GAT-3/GAT-4) function. To counteract the heightened inhibition, we administered *in vivo* a benzodiazepine inverse agonist specific for α5-subunit-containing extrasynaptic GABA_A receptors at a delay after stroke. This treatment produced an early and sustained recovery of motor function. Genetically lowering the number of α5- or δ-subunit-containing GABA_A receptors responsible for tonic inhibition also proved beneficial for recovery after stroke, consistent with the therapeutic potential of diminishing extrasynaptic GABA_A receptor function. Together, our results identify new pharmacological targets and provide the rationale for a novel strategy to promote recovery after stroke and possibly other brain injuries.

Stroke is a major source of disability, confining one-third of stroke survivors to nursing homes or institutional settings⁴. Recent studies have shown that the brain has a limited capacity for repair after stroke. Neural repair after stroke involves re-mapping of cognitive functions in tissue adjacent to or connected with the stroke^{5,6}. Functional recovery in this peri-infarct tissue involves changes in neuronal excitability that alter the brain's representation of motor and sensory functions. Stimulation of peri-infarct cortex enhances local neuronal excitability through a process that involves long-term potentiation, alters sensorimotor maps and improves the use of affected limbs^{5–8}. The inhibitory neurotransmitter GABA is critical for cortical plasticity and sensory mapping. Altering GABAergic transmission changes sensory maps during the critical period of cortical development⁹ and produces rapid alterations in adult cortical maps that resemble changes occurring after stroke^{10,11}. Alterations in cortical maps through blockade of GABAergic signalling are associated with fundamental changes in cellular excitability including long-term potentiation¹². In a similar manner to normal cortical plasticity, GABAergic mechanisms may mediate changes in neuronal excitability that have a central role in functional recovery of peri-infarct cortex after stroke.

Cortical GABAergic signalling through GABA_A receptors is divided into synaptic (phasic) and extrasynaptic (tonic) components. Tonic active extrasynaptic GABA_A receptors set an excitability threshold for neurons^{13,14}. Extrasynaptic GABA_A receptors primarily consist of α5- or δ-subunit-containing receptors^{13,14}. Pharmacological and genetic knockdown of α5-GABA_A receptors enhance long-term potentiation and improve performance on learning and memory tasks^{15,16}.

The selective effects of extrasynaptic GABA_A receptors on cellular excitability and plasticity, and the evidence that changes in neuronal excitability underlie functional reorganization in peri-infarct cortex, indicate that this system may have a role in recovery after stroke. We find that stroke increases tonic GABAergic transmission in peri-infarct cortex and that dampening this tonic inhibition produces an early and robust gain of motor recovery after stroke (Supplementary Fig. 1, schematic summary).

We examined neuronal excitability in the peri-infarct cortex of mice during the period of recovery and reorganization after a photothrombotic stroke to forelimb motor cortex. Whole-cell voltage-clamp recordings in *in vitro* brain slices prepared at 3, 7 and 14 days after stroke (Fig. 1a) showed a significant increase in GABA_A-receptor-mediated tonic inhibition (I_{tonic}) in layer 2/3 pyramidal neurons, compared to neurons from sham controls (control: 8.05 ± 0.80 pA pF⁻¹, $n = 24$, versus after stroke: 13.6 ± 1.41 pA pF⁻¹, $n = 45$, Mann-Whitney U -test, $P < 0.05$; Fig. 1b–d). I_{tonic} remained raised from 3 to 14 days after stroke (Supplementary Fig. 2a). The mean phasic excitation remained unchanged over the 2-week period after stroke

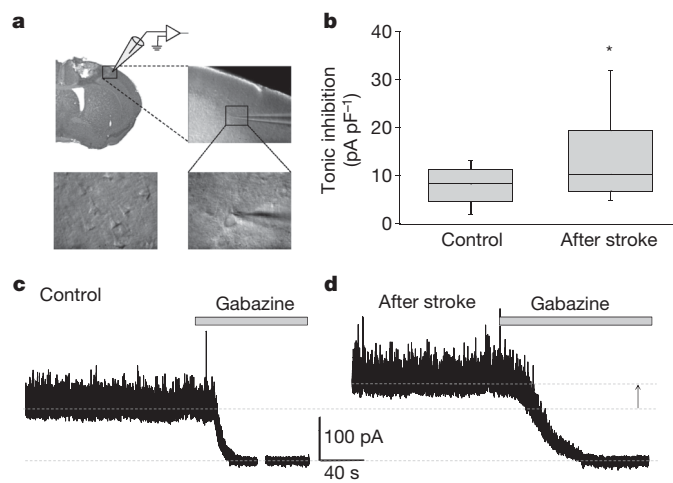


Figure 1 | Increased tonic inhibition in peri-infarct cortex. **a**, Images showing the peri-infarct recording site. Whole-cell patch-clamp recordings were made from post-stroke brain slices, within 200 μm of the infarct (top left; image under 2× magnification), from layer 2/3 (top right; 10×) pyramidal neurons (bottom panels; 40×). **b**, Box plot (boxes, 25–75%; whiskers, 10–90%; lines, median) showing significantly increased tonic inhibition in peri-infarct cortex (asterisk, $P < 0.05$; see Supplementary Fig. 2 for further analyses). **c**, **d**, Representative traces showing tonic inhibitory currents in control and peri-infarct neurons, respectively. Tonic currents were revealed by the shift in holding currents after blocking all GABA_A receptors with gabazine (> 100 μM). Cells were voltage clamped at +10 mV.

¹Department of Neurology, The David Geffen School of Medicine at UCLA, 635 Charles Young Drive South, Los Angeles, California 90095, USA. ²Interdepartmental PhD Program for Neuroscience, The David Geffen School of Medicine at UCLA, 635 Charles Young Drive South, Los Angeles, California 90095, USA. ³Department of Physiology, The David Geffen School of Medicine at UCLA, 635 Charles Young Drive South, Los Angeles, California 90095, USA. [†]Present address: Departments of Psychology and Anatomy and Structural Biology, University of Otago, PO Box 913, Dunedin 9013, New Zealand.

*These authors contributed equally to this work.

(Supplementary Fig. 3a, c). The mean phasic inhibition was unchanged except for a transient decrease at 7 days after stroke (Supplementary Fig. 3b, d). The resting membrane and GABA reversal potentials were both unchanged (Supplementary Fig. 3e, f).

Tonic inhibition is effectively controlled by the degree of extracellular GABA uptake through neuronal and astrocytic GATs¹⁴. We applied a GAT-1-selective antagonist, NO-711 (10 μ M), and found a significantly greater effect (percentage increase in I_{tonic} after GAT blockade) in post-stroke neurons ($94.0 \pm 16.3\%$, $n = 10$) than in controls ($34.3 \pm 11.4\%$, $n = 6$; $P < 0.05$; Fig. 2a). Co-application of NO-711 and the GAT-3/GAT-4-selective antagonist SNAP-5114 (40 μ M) produced a substantial increase in I_{tonic} in controls ($300.6 \pm 46.0\%$, $n = 4$; Fig. 2a), revealing the synergistic actions of GATs in the cortex as previously proposed¹⁷. In post-stroke neurons, co-application only produced an effect ($110.7 \pm 32.0\%$, $n = 5$) similar to GAT-1 blockade alone ($P = 0.68$; Fig. 2a), indicating a dysfunction in GAT-3/GAT-4 after stroke. Sequential blockade of the two GATs confirmed the impairment after stroke, as peri-infarct I_{tonic} showed no further response to GAT-3/GAT-4 blockade after the initial GAT-1 block, in contrast to responses shown in controls (Fig. 2b, c). This effect was not due to receptor saturation, as I_{tonic} showed a further increase in response to a raised concentration of GABA under GAT blockade (Supplementary Fig. 4a). Western-blot analysis confirmed a reduced GAT-4 level in peri-infarct cortex, whereas GAT-1 levels were unchanged (Supplementary Fig. 5).

We proposed that the chronically raised tonic inhibition in the peri-infarct region may antagonize the neuronal plasticity required for functional recovery after stroke. Therefore, we tested whether reducing the excessive tonic inhibition would improve functional recovery. Of the two GABA_A receptor subtypes shown to underlie tonic inhibition in cortical neurons, $\alpha 5$ -GABA_A receptors can be antagonized specifically by L655,708, a benzodiazepine inverse agonist¹⁶, whereas no specific antagonist exists for δ -GABA_A receptors. L655,708 (100 nM) decreased I_{tonic} in control neurons by $-13.3 \pm 5.2\%$ ($n = 4$), but produced a significantly greater decrease in post-stroke neurons ($-30.0 \pm 4.1\%$, $n = 13$; $P < 0.05$; Fig. 2d, e), which reverted I_{tonic} back to control level (control: see earlier versus after stroke plus L655,708: 140.8 ± 18.5 pA, $n = 13$; $P = 0.702$; Fig. 2f). L655,708 produced only minimal effects on phasic inhibitory currents in both post-stroke and control conditions (Supplementary Fig. 4b).

Next we tested the effects of reducing tonic inhibition on functional recovery after stroke, using measures of forelimb and hindlimb motor control. Stroke produced an increase in the number of foot faults in the grid-walking task, and a decrease in forelimb symmetry in the cylinder task from 7 days after stroke. Chronic treatment with L655,708 starting 3 days after stroke resulted in a dose-dependent maximal gain of function beginning from 7 days after stroke in both tasks ($P < 0.001$; Fig. 3a–c). Acute treatment with L655,708 just before behavioural testing had a minimal effect on stroke recovery (Supplementary Fig. 7). To assess the necessity of long-term administration, we discontinued

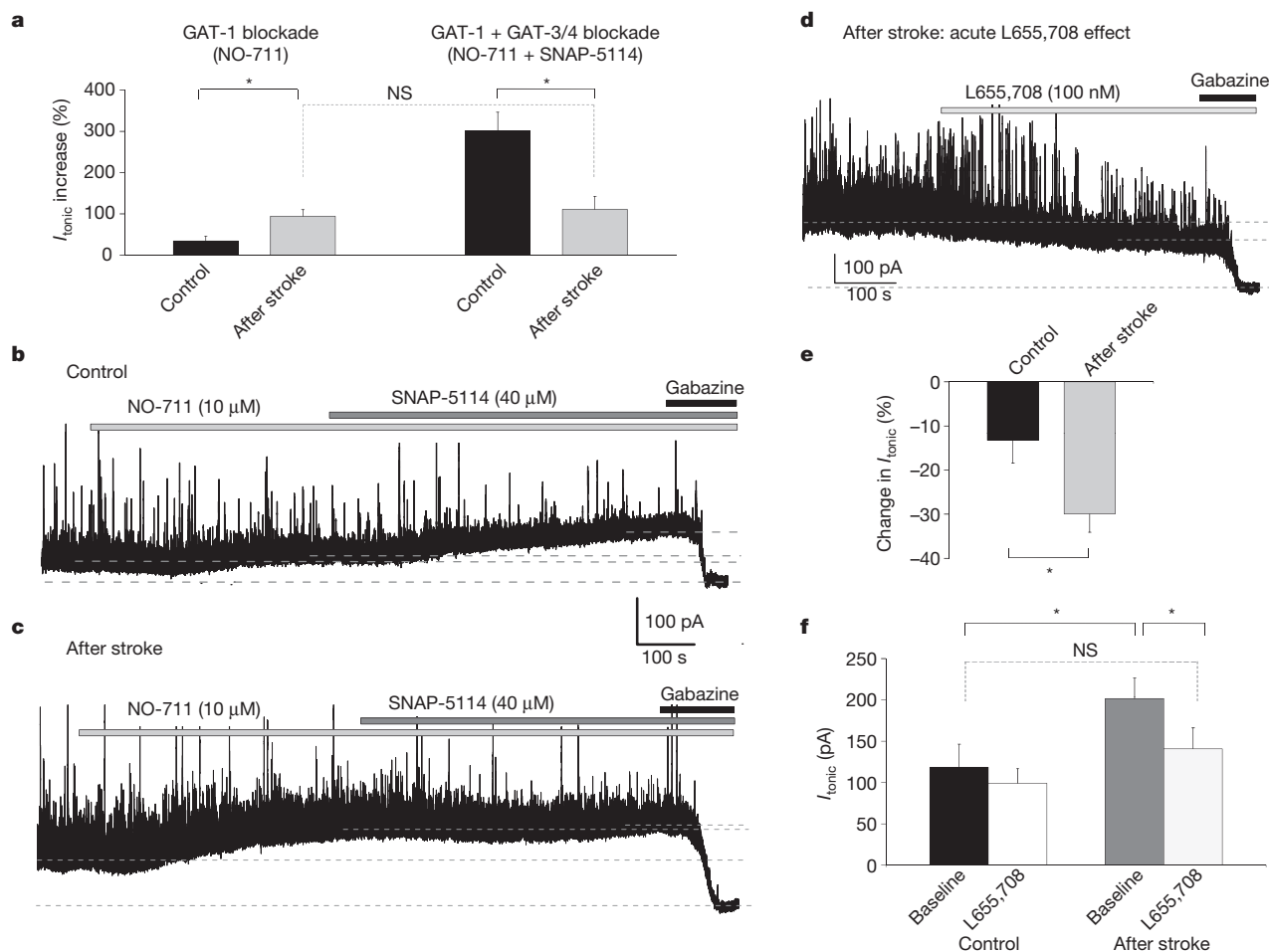


Figure 2 | Impairment in GABA transport and effect of blocking $\alpha 5$ -GABA_A receptors after stroke. **a**, Blocking GAT-1 (NO-711) produced a higher percentage increase in I_{tonic} after stroke; combined blockade of GAT-1 and GAT-3/GAT-4 (NO-711 plus SNAP-5114) produced a substantial I_{tonic} increase in controls but only an increase equivalent to blocking GAT-1 alone

after stroke. **b**, **c**, I_{tonic} in sequential drug applications. Note the lack of response to SNAP-5114 application in the post-stroke slice. **d**, L655,708 reduced I_{tonic} . **e**, L655,708 significantly decreased I_{tonic} after stroke. **f**, Drug treatment reverted post-stroke I_{tonic} to near-control level. *, $P < 0.05$; NS, not significant; bar graphs represent mean \pm s.e.m.

L655,708 treatment after 2 weeks and found a decrease in functional gains, although these mice still performed better than vehicle-treated stroke controls (Supplementary Fig. 6).

To corroborate further the role of reduced tonic inhibition in enhancing stroke recovery, we tested mice with deletions of either $\alpha 5$ - or δ -subunit-containing GABA_A receptors (Gabra5^{-/-} and Gabrd^{-/-}; Methods)¹⁸. Gabra5^{-/-} animals showed significantly better motor recovery after stroke, comparable to L655,708-treated wild-type animals (Fig. 3d–f). In addition, Gabra5^{-/-} animals showed a significant reduction in hindlimb foot faults (Fig. 3e). Gabrd^{-/-} animals also showed significant improvements in motor recovery (Fig. 3d–f), but to a lesser extent than the Gabra5^{-/-} mice. Thus, modulation of $\alpha 5$ -GABA_A receptors affords greater functional gains in motor recovery than δ -GABA_A receptors, and genetic removal of $\alpha 5$ -GABA_A receptors produces a more widespread increase in motor recovery than pharmacological antagonism. Administration of L655,708 to Gabrd^{-/-} mice produced an even greater recovery, confirming the beneficial effect of reducing peri-infarct tonic inhibition (Fig. 3a, c).

Low/sub-seizure dosing of picrotoxin (PTX; 0.1 mg kg⁻¹, administered intraperitoneally (i.p.)), a use-dependent GABA_A receptor antagonist, enhances learning and memory in transgenic mouse models of Alzheimer's and other cognitive impairments by reversing

an increased GABAergic inhibitory tone, acting at both synaptic and extrasynaptic GABA_A receptors^{19,20}. The pharmacological effects of PTX on reducing phasic and tonic inhibition were not altered after stroke (Supplementary Table 2). PTX given to animals from 3 days after stroke resulted in a significant gain of forelimb function on the grid-walking task compared to vehicle-treated stroke controls ($P < 0.05$; Supplementary Fig. 9a). No significant changes were observed in hindlimb function or forelimb asymmetry (Supplementary Fig. 9b, c). Combined L655,708 and PTX treatment showed similar initial functional gains compared to stroke and L655,708 alone; however, prolonged PTX and L655,708 treatment produced a deterioration in motor function such that the performance progressively worsened at late periods after stroke (Supplementary Fig. 9). These data indicate that increasing cortical excitability too far or reducing phasic inhibition negatively affects functional recovery.

An important element in stroke treatment is the timing of drug delivery. GABA_A receptor agonists administered at the time of stroke decrease stroke size²⁰. Therefore, dampening tonic inhibition too early after stroke may produce an opposite effect, that is, increased cell death. To test this, we assessed stroke volume at 7 days after stroke, in animals treated with 1) vehicle; 2) L655,708 from stroke onset; and 3) L655,708 from day 3 after stroke. Stroke volumes were similar between mice

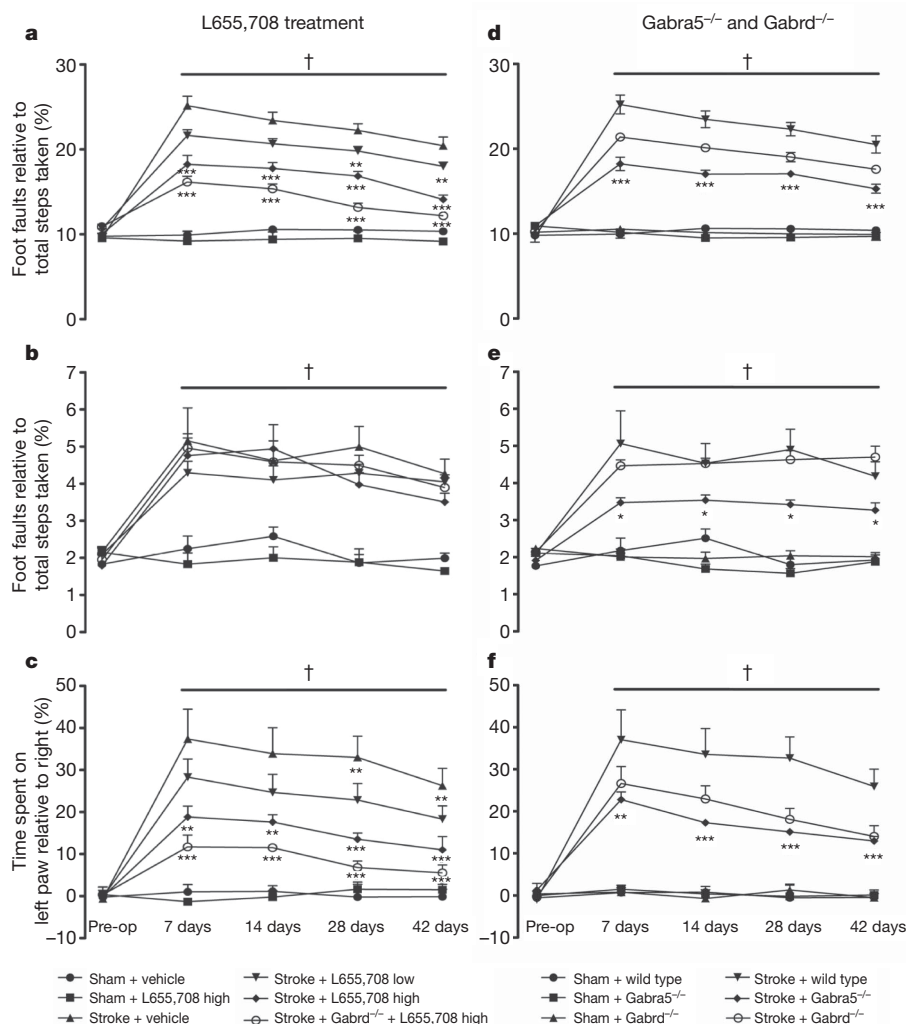


Figure 3 | Behavioural recovery after stroke with L655,708 treatment and in Gabra5^{-/-} and Gabrd^{-/-} animals. a–c, L655,708 treatment starting from 3 days after stroke resulted in a dose-dependent improvement in functional recovery after stroke. d–f, Gabra5^{-/-} and Gabrd^{-/-} mice also showed decreased motor deficits after stroke. Functional recovery was assessed with

forelimb (a, d) and hindlimb foot faults (b, e), and on forelimb asymmetry (c, f). Low-dose L655,708 = 200 $\mu\text{g kg}^{-1} \text{ day}^{-1}$ per animal; high-dose L655,708 = 400 $\mu\text{g kg}^{-1} \text{ day}^{-1}$ per animal. Pre-op, pre-operation. Data are mean \pm s.e.m. †, $P \leq 0.001$ stroke plus vehicle versus sham; *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$ versus stroke plus vehicle.

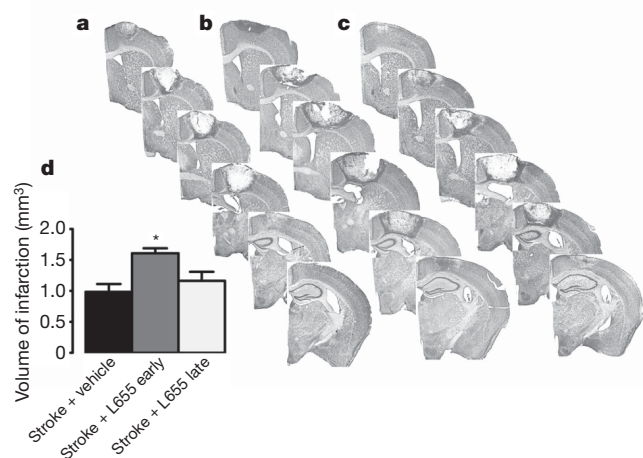


Figure 4 | Inflection point in L655,708 treatment effect on infarct size. **a–d**, Representative Nissl-stained sections at 7 days after stroke from stroke plus vehicle treatment (**a**), stroke plus L655,708 treatment starting at the time of stroke (**b**) and stroke plus L655,708-treatment starting from 3 days post-insult (**c**). Quantification of the stroke volume is shown in panel **d**. Data are mean \pm s.e.m. for $n = 4$ per group; * = $P \leq 0.05$.

treated with vehicle and L655,708 from day 3 (Fig. 4). In contrast, stroke volume was significantly increased in animals treated with L655,708 from stroke onset ($P < 0.05$; Fig. 4). These data indicate a critical timeframe for therapeutically dampening tonic inhibition after stroke: reduction too early would exacerbate stroke damage, whereas delaying treatment by 3 days would promote functional recovery without altering stroke size. Genetic deletion of $\alpha 5$ - or δ -GABA_A receptors did not affect infarct size or neuronal number in peri-infarct cortex (Supplementary Fig. 8). Unlike pharmacological antagonism of $\alpha 5$ -GABA_A-receptor-mediated inhibition, in *Gabra5*^{−/−} and *Gabrd*^{−/−} mice, the genomic absence of one of the extrasynaptic GABA_A receptors may trigger compensatory upregulation of the other receptor¹³, thus obscuring their roles in neuroprotection immediately after stroke.

Current therapies that promote functional recovery after stroke are limited to physical rehabilitation⁴. Here, by identifying an excessive tonic inhibition after stroke, we have found promising new targets for pharmacological interventions to promote recovery. The increase in tonic inhibition in cortical pyramidal neurons occurs during precisely the same time period important for cortical map plasticity and recovery^{1–3}. Alterations in other aspects of cortical signalling have also been described during this period, including altered GABA_A receptor subunits, glutamate receptor expression and neuronal network properties^{21–24}. Protein levels of GAT-1 were shown to be decreased in peri-infarct cortex in some rodent stroke models, and reactive astrocytes show reduced uptake of other neurotransmitters²⁴. However, there are conflicting data on GABA_A receptor levels after stroke^{23–26}. We found a decreased protein level and compromised function of GAT-3/GAT-4 in peri-infarct cortex. The increase in tonic inhibition may curtail cortical plasticity and spontaneous recovery after stroke, and is consistent with tonic GABAergic inhibition exerting a causal role in limiting motor recovery in stroke.

Non-selectively decreasing GABAergic tone facilitates neuronal plasticity in genetic models of cognitive diseases^{19,20}. We show for the first time, to our knowledge, that antagonizing a raised tonic inhibition enhances motor recovery after stroke, consistent with the idea that molecular and cellular events of neuronal plasticity are dampened in the peri-infarct zone, and promoting this plasticity facilitates functional recovery. Together, our results have identified novel pharmacological targets and provide a rational basis for developing future therapies to promote recovery after stroke and possibly other brain injuries.

METHODS SUMMARY

Photothrombotic model of focal stroke. Focal stroke was induced by photothrombosis in adult male C57BL/6 mice (aged 2–4 months) as described previously²⁷.

Slice preparation for electrophysiology. Following decapitation, brains were rapidly removed and placed into an *N*-methyl-D-glucamine (NMDG)-based cutting solution to enhance neuronal viability²⁸. Coronal slices (350 μ m) were cut and transferred to an interface-style chamber containing artificial cerebrospinal fluid (ACSF) as previously described¹³. Recordings were made from intact peri-infarct cortical layer-2/3 pyramidal neurons and analysed as previously described¹³.

In vivo drug administration. L655,708 was dissolved in dimethylsulphoxide (DMSO) and then diluted 1:1 in 0.9% saline. L655,708-filled ALZET-1002 pumps were implanted at 3 days after stroke and replaced every two weeks. In acute administration studies, 5 mg kg^{−1} L655,708 was administered i.p. 30 min before testing. The concentration in one minipump, 5 mM, delivers a 200 μ g kg^{−1} day^{−1} dose in mice. With one or two minipumps implanted, this provides a dose escalation. PTX (0.1 mg kg^{−1} i.p. twice a day) starting 3 days after stroke was administered alone or together with L655,708.

Behavioural analysis. Mice were videotaped during walking and exploratory behaviour in the grid-walking and cylinder/rearing tasks, and were tested at approximately the same time each day during the nocturnal period²⁹. Baseline behavioural measurements were obtained one week before surgery. Post-stroke animals were assessed at weeks 1, 2, 4 and 6.

Infarct-size measurement. For the histological assessment of infarct size, brains were processed at 7 days after stroke using cresyl violet as previously described³⁰.

Statistical analyses. All data are expressed as mean \pm s.e.m. For electrophysiological comparisons between control and post-stroke animals, the Mann–Whitney non-parametric test was used. For multiple comparisons across days after stroke, one-way analysis of variance (ANOVA) and Newman–Keuls' multiple pair-wise comparisons for post-hoc comparisons were used. For behavioural testing, differences between treatment groups were analysed using two-way ANOVA with repeated measures and Newman–Keuls' multiple pair-wise comparisons. The level of significance was set at $P < 0.05$.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 November 2009; accepted 16 September 2010.

Published online 3 November 2010.

- Cramer, S. C. Repairing the human brain after stroke: I. Mechanisms of spontaneous recovery. *Ann. Neurol.* **63**, 272–287 (2008).
- Brown, C. E., Aminoltejeri, K., Erb, H., Winship, I. R. & Murphy, T. H. *In vivo* voltage-sensitive dye imaging in adult mice reveals that somatosensory maps lost to stroke are replaced over weeks by new structural and functional circuits with prolonged modes of activation within both the peri-infarct zone and distant sites. *J. Neurosci.* **29**, 1719–1734 (2009).
- Dijkhuizen, R. M. et al. Correlation between brain reorganization, ischemic damage, and neurologic status after transient focal cerebral ischemia in rats: a functional magnetic resonance imaging study. *J. Neurosci.* **23**, 510–517 (2003).
- Dobkin, B. H. Training and exercise to drive poststroke recovery. *Nature Clin. Pract. Neurol.* **4**, 76–85 (2008).
- Carmichael, S. T. Cellular and molecular mechanisms of neural repair after stroke: making waves. *Ann. Neurol.* **59**, 735–742 (2006).
- Nudo, R. J. Mechanisms for recovery of motor function following cortical damage. *Curr. Opin. Neurobiol.* **16**, 638–644 (2006).
- Alonso-Alonso, M., Fregni, F. & Pascual-Leone, A. Brain stimulation in poststroke rehabilitation. *Cerebrovasc. Dis.* **24** (suppl. 1), 157–166 (2007).
- Di Lazzaro, V. et al. Motor cortex plasticity predicts recovery in acute stroke. *Cereb. Cortex* **20**, 1523–1528 (2010).
- Hensch, T. K. Critical period plasticity in local cortical circuits. *Nature Rev. Neurosci.* **6**, 877–888 (2005).
- Donoghue, J. P., Suner, S. & Sanes, J. N. Dynamic organization of primary motor cortex output to target muscles in adult rats. II. Rapid reorganization following motor nerve lesions. *Exp. Brain Res.* **79**, 492–503 (1990).
- Foeller, E., Celikel, T. & Feldman, D. E. Inhibitory sharpening of receptive fields contributes to whisker map plasticity in rat somatosensory cortex. *J. Neurophysiol.* **94**, 4387–4400 (2005).
- Hess, G., Aizenman, C. D. & Donoghue, J. P. Conditions for the induction of long-term potentiation in layer II/III horizontal connections of the rat motor cortex. *J. Neurophysiol.* **75**, 1765–1778 (1996).
- Glykys, J. & Mody, I. Hippocampal network hyperactivity after selective reduction of tonic inhibition in GABA_A receptor $\alpha 5$ subunit-deficient mice. *J. Neurophysiol.* **95**, 2796–2807 (2006).
- Walker, M. C. & Semyanov, A. Regulation of excitability by extrasynaptic GABA(A) receptors. *Results Probl. Cell Differ.* **44**, 29–48 (2008).
- Collinson, N. et al. Enhanced learning and memory and altered GABAergic synaptic transmission in mice lacking the $\alpha 5$ subunit of the GABA_A receptor. *J. Neurosci.* **22**, 5572–5580 (2002).

16. Atack, J. R. *et al.* L-655,708 enhances cognition in rats but is not proconvulsant at a dose selective for $\alpha 5$ -containing GABA_A receptors. *Neuropharmacology* **51**, 1023–1029 (2006).
17. Keros, S. & Hablitz, J. J. Subtype-specific GABA transporter antagonists synergistically modulate phasic and tonic GABA_A conductances in rat neocortex. *J. Neurophysiol.* **94**, 2073–2085 (2005).
18. Glykys, J. & Mody, I. Activation of GABA_A receptors: views from outside the synaptic cleft. *Neuron* **56**, 763–770 (2007).
19. Yoshiike, Y. *et al.* GABA_A receptor-mediated acceleration of aging-associated memory decline in APP/PS1 mice and its pharmacological treatment by picrotoxin. *PLoS ONE* **3**, e3029 (2008).
20. Cui, Y. *et al.* Neurofibromin regulation of ERK signaling modulates GABA release and learning. *Cell* **135**, 549–560 (2008).
21. Ginsberg, M. D. Neuroprotection for ischemic stroke: past, present and future. *Neuropharmacology* **55**, 363–389 (2008).
22. Que, M. *et al.* Changes in GABA(A) and GABA(B) receptor binding following cortical photothrombosis: a quantitative receptor autoradiographic study. *Neurosci. Lett.* **93**, 1233–1240 (1999).
23. Redecker, C., Luhmann, H. J., Hagemann, G., Fritschy, J. M. & Witte, O. W. Differential downregulation of GABA_A receptor subunits in widespread brain regions in the freeze-lesion model of focal cortical malformations. *J. Neurosci.* **20**, 5045–5053 (2000).
24. Frahm, C. *et al.* Regulation of GABA transporter mRNA and protein after photothrombotic infarct in rat brain. *J. Comp. Neurol.* **478**, 176–188 (2004).
25. Neumann-Haefelin, T. *et al.* Immunohistochemical evidence for dysregulation of the GABAergic system ipsilateral to photochemically induced cortical infarcts in rats. *Neuroscience* **87**, 871–879 (1998).
26. Kharlamov, E. A., Downey, K. L., Jukkola, P. I., Grayson, D. R. & Kelly, K. M. Expression of GABA_A receptor $\alpha 1$ subunit mRNA and protein in rat neocortex following photothrombotic infarction. *Brain Res.* **1210**, 29–38 (2008).
27. Lee, J. K., Kim, J. E., Sivula, M. & Strittmatter, S. M. Nogo receptor antagonism promotes stroke recovery by enhancing axonal plasticity. *J. Neurosci.* **24**, 6209–6217 (2004).
28. Tanaka, Y., Furuta, T., Yanagawa, Y. & Kaneko, T. The effects of cutting solutions on the viability of GABAergic interneurons in cerebral cortical slices of adult mice. *J. Neurosci. Methods* **171**, 118–125 (2008).
29. Baskin, Y. K., Dietrich, W. D. & Green, E. J. Two effective behavioral tasks for evaluating sensorimotor dysfunction following traumatic brain injury in mice. *J. Neurosci. Methods* **129**, 87–93 (2003).
30. Ohab, J. J., Fleming, S., Blesch, A. & Carmichael, S. T. A neurovascular niche for neurogenesis after stroke. *J. Neurosci.* **26**, 13007–13016 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements I.M., A.N.C. and S.T.C. were supported by The Dr. Miriam and Sheldon G. Adelson Medical Research Foundation. S.T.C. was supported by the Larry L. Hillblom Foundation, I.M. was supported by the Coelho Endowment and National Institutes of Health/National Institute of Neurological Disorders and Stroke grant NS30549. This manuscript was completed partially during tenure of an American Heart Association Postdoctoral Fellowship, a Repatriation Fellowship from the New Zealand Neurological Foundation and the Sir Charles Hercus Fellowship from the Health Research Council of New Zealand (A.N.C.). We thank E. O. Mann, J. Chu, J. J. Overman, J. Zhong and R. M. Lazaro for discussion and assistance.

Author Contributions A.N.C. performed the behavioural, histological and immunohistochemical studies; B.S.H. carried out the electrophysiological experiments; and S.E.M. performed the immunohistochemical and western blot work. A.N.C., B.S.H., I.M. and S.T.C. designed the experiments, analysed data, prepared figures and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.T.C. (scarmichael@mednet.ucla.edu).

METHODS

Photothrombotic model of focal stroke. Under isoflurane anaesthesia (2–2.5% in a 70% N₂O/30% O₂ mixture), 2–4 month-old adult C57BL/6 (Charles River) male mice were placed in a stereotaxic apparatus, the skull exposed through a midline incision, cleared of connective tissue and dried. A cold light source (KL1500 LCD, Zeiss) attached to a 40× objective giving a 2-mm diameter illumination was positioned 1.5 mm lateral from Bregma, and 0.2 ml of Rose Bengal solution (Sigma; 10 g l⁻¹ in normal saline, i.p.) was administered³¹. After 5 min, the brain was illuminated through the intact skull for 15 min. Rose Bengal produces singlet oxygen under light excitation, which damages and occludes vascular endothelium, resulting in focal cortical stroke under the region of illumination (Fig. 4), circumscribed by peri-infarct tissue with normal neuronal cell number (Supplementary Fig. 8). Two-to-four-month-old adult male Gabra5^{-/-} and Gabrd^{-/-} mice³² received a stroke as described earlier. These mice had been backcrossed to C57BL/6 mice in excess of 15 generations, and were compared in behavioural studies to wild-type C57BL/6. Body temperature was maintained at 36.9 ± 0.4 °C with a heating pad throughout the operation and did not vary by drug or genetic condition. This stroke method produces a small stroke in the mouse forelimb region of the motor cortex (Fig. 4). Sample size was ten per group for Gabra5^{-/-} and Gabrd^{-/-} in stroke/behavioural studies. Sample size was eight per group for each condition in dosing of L655,708 (Fig. 3).

Blood pressure (systolic and diastolic) and heart rate were measured in separate cohorts of wild-type (C57BL/6) mice, with or without L655,708 administration via ALZET minipumps from 3 days after stroke, and in Gabra5^{-/-} and Gabrd^{-/-} mice, before, during and after stroke, using a standard non-invasive tail-cuff method (Coda). There were no significant differences in heart rate or blood pressure by treatment or genotype (Supplementary Table 1). All studies in this manuscript complied with the Stroke Therapy Academic Industry Roundtable (STAIR) criteria^{33,34} for stroke investigations in measuring physiological parameters, monitoring treatment effects for at least one month, analysing treatment effects blinded to conditions, using dose–response studies and use of a drug administration route with blood–brain-barrier penetration.

Whole-cell voltage-clamp electrophysiology. Slices were submerged in the recording chamber and continuously perfused (5–8 ml min⁻¹) with oxygenated ACSF (32–34 °C). Visualized patch-clamp recordings from layer-2/3 pyramidal neurons were performed at 40× using infrared oblique-illumination (Leica DM-LFS; Hamamatsu CCD camera C3077-78).

Control recordings were made from cells of sham-operated animals at similar locations as those recorded in post-stroke animals. Microelectrodes (3–5 MΩ) were filled with a cesium-methylsulphonate (CsMeSO₄)-based internal pipette solution, containing 120 mM CsMeSO₄, 10 mM CsCl, 5 mM TEA-Cl, 1.5 mM MgCl₂, 10 mM HEPES, 0.1 mM EGTA, 2 mM Na-ATP, 0.5 mM Na-GTP and 5 mM QX-314, pH 7.25–7.30 with CsOH, 275–285 mOsm. The recording ACSF was supplemented with 5 μM GABA to replenish the extracellular GABA concentration reduced by the high-flow perfusion of the slices³². For recording *I*_{tonic} at -70 mV, a high-CsCl-based internal solution was used, containing 140 mM CsCl, 1 mM MgCl₂, 10 mM HEPES, 0.1 mM EGTA, 4 mM NaCl, 2 mM Mg-ATP, 0.3 mM Na-GTP and 5 mM QX-314, pH ~7.3, ~275 mOsm l⁻¹, with ACSF containing 3 mM kynurenic acid to block glutamatergic currents.

Neurons were voltage clamped in whole-cell configuration using a MultiClamp-700A amplifier (Molecular Devices); all recordings were low-pass-filtered at 3 kHz (8-pole Bessel) and digitized online at 10 kHz (National Instruments PCI-MIO-16E-4 board). Series resistance and whole-cell capacitance were estimated from fast transients evoked by a 5-mV step and compensated to 75%. Excitatory postsynaptic currents (EPSCs) and inhibitory postsynaptic currents (IPSCs) were recorded by voltage clamping sequentially at -70 mV and +10 mV, respectively.

All drugs were purchased from Sigma or Tocris. L-655,708 and SNAP-5114 were dissolved in DMSO then diluted 1:1,000 in H₂O. NO-711, gabazine and GABA were dissolved in H₂O.

Tonic inhibitory current and mean phasic current determination. Custom-written macros running under Igor Pro v.6.0 (WaveMetrics) were used to analyse the digitized recordings to determine the values of tonic currents and mean phasic currents, as previously described³¹. *I*_{tonic} was recorded as the reduction in baseline holding currents (*I*_{hold}) after bath-applying a saturating amount (>100 μM) of the GABA_A-receptor antagonist SR-95531 (gabazine), while voltage clamping at +10 mV. NO-711, SNAP-5114 and L-655,708 were added to the recording ACSF via perfusion and their effects on *I*_{tonic} were recorded as the post-drug shift in *I*_{hold}. Drug perfusion was continued until the shifting *I*_{hold} remained steady for 1–2 min.

To determine the mean phasic current (*I*_{mean}), a 60-s segment containing either EPSCs or IPSCs was selected, and an all-point histogram was plotted for every 10,000 points (every 1 s), smoothed, and fitted with a Gaussian to obtain the mean baseline current. All baseline mean values were then plotted and linear trends subtracted to normalize the mean baseline current to 0 pA. After baseline normalization, the values

of each 10,000 points (each 1 s) were averaged to yield the value of *I*_{mean} (in pA s⁻¹) for each 1-s epoch. The averaged *I*_{mean} value of a 60-s segment was reported as the phasic *I*_{mean} value for either the spontaneous EPSC or IPSC. Synaptic event kinetics (that is, frequency, peak amplitude, 10–90% rise time and weighted decay time constant) are analysed by custom-written LabView-based software (EVAN), as previously described³². For comparison of the IPSC peak amplitudes under control and PTX-treated conditions (Supplementary Table 2), the largest-amplitude count-matched method³² was used, whereby the amplitude values in a given recording were sorted and the largest *x* number of events under control conditions were averaged and compared to the average of an equally matched *x* number of events recorded during a similar time period under the PTX condition, with *x* being the total number of events detected under the 10 μM PTX condition. This method circumvents the erroneous comparison of average amplitudes when considering the effects of a receptor antagonist that reduces the smaller events (in control conditions) down to the noise level.

Measurements of neuronal resting membrane potential and GABA reversal potential. To estimate neuronal resting membrane potential (*V*_{rest}), the cell-attached recording technique^{35,36} was used. Briefly, depolarizing voltage ramps (-100 to +200 mV) were applied to cell-attached patches to activate voltage-gated K⁺ channels and establish the K⁺ current reversal potential, which provides a measure of the *V*_{rest}, given near equimolar K⁺ inside the cell and the pipette. The GABA reversal potential (*E*_{GABA}) was estimated by measuring the K⁺ reversal potential after activating GABA_A receptors with 50 μM muscimol. Recordings were made using a solution containing the following: 135 mM K⁺ gluconate, 5 mM KCl, 2 mM MgCl₂, 10 mM HEPES, 0.1 mM EGTA, 4 mM Na-ATP, 0.3 mM Na-GTP, pH 7.3, 273 mOsm l⁻¹. A junction potential of 9 mV was measured and then subtracted from voltage values of all measurements.

Fitting of multiple distributions to cumulative probability plots. The fitting of multiple distributions to a cumulative probability plot (Supplementary Fig. 2) was done as follows. Cumulative probabilities of the variable *x* (that is, *P*(*x*)) were calculated and fitted by one or more normal curves approximated by the logistic equation³⁷:

$$P(x) = \sum_{i=1}^n R_i \frac{x^{p_i}}{x^{p_i} + \bar{x}_i^{p_i}}$$

where *R*₁...*R*_{*n*} are the ratios of the *n* normal distributions (such as $\sum_{i=1}^n R_i = 1$), \bar{x}_1 ... \bar{x}_n are the individual means, and *p*₁...*p*_{*n*} are steepness factors related to the *n* standard deviations (s.d.₁...s.d._{*n*}).

Grid-walking task. The grid-walking apparatus was manufactured as previously described^{38,39}, using a 12-mm square wire mesh with a grid area of 32 cm/20 cm/50 cm (length/width/height). A mirror was placed beneath the apparatus to allow video footage in order to assess the animals' stepping errors (foot faults). Each mouse was placed individually on top of the elevated wire grid and allowed to freely walk for a period of 5 min. Video footage was analysed offline by raters blind to the treatment groups. The total number of foot faults for each limb, along with the total number of non-foot-fault steps, was counted, and a ratio between foot faults and total steps taken calculated. Per cent foot faults were calculated by: number of foot faults / (foot faults + number of non-foot-fault steps) × 100. A ratio between foot faults and total steps taken was used to take into account differences in the degree of locomotion between animals and trials. A step was considered a foot fault if it was not providing support and the foot went through the grid hole. Furthermore, if an animal was resting with the grid at the level of the wrist, this was also considered a fault. If the grid was anywhere forward of the wrist area then this was considered as a normal step.

Spontaneous forelimb task (cylinder task). The spontaneous forelimb task encourages the use of forelimbs for vertical wall exploration/press in a cylinder^{38,39}. When placed in a cylinder, the animal rears to a standing position, while supporting its weight with either one or both of its forelimbs on the side of the cylinder wall. Mice were placed inside a Plexiglas cylinder (15 cm in height with a diameter of 10 cm) and videotaped for 5 min. Videotaped footage of animals in the cylinder was evaluated quantitatively in order to determine forelimb preference during vertical exploratory movements. While the video footage was played in slow motion (1/5th real time speed), the time (seconds) during each rear that each animal spent on either the right forelimb, the left forelimb, or on both forelimbs were calculated. Only rears in which both forelimbs could be clearly seen were timed. The percentage of time spent on each limb was calculated and these data were used to derive an SFL asymmetry index ((per cent ipsilateral use) – (per cent contralateral use)). The 'contact time' method of examining the behaviour was chosen over the 'contact placement' method, as described in ref. 38, as it takes into account the slips that often occur during a bilateral wall press after photothrombosis.

Western blot. Seven days after stroke mice were decapitated, the brains rapidly removed and peri-infarct cortex microdissected and frozen (*n* = 5). The equivalent region of cortex was taken in control, non-operated mice (*n* = 3). Samples were homogenized in radioimmunoprecipitation (RIPA) buffer (Pierce) and centrifuged at 20,000g at 4 °C for 10 min. Supernatant was collected as protein extract

and stored at -80°C . Western blot was performed as described⁴⁰. One-hundred micrograms of protein from each sample was diluted in $7.5\ \mu\text{l}$ of $2\times$ SDS-sample buffer gel (Invitrogen) containing dithiothreitol (DTT) (Sigma) and brought to a final volume of $15\ \mu\text{l}$ with RIPA buffer. Samples were denatured at 95°C , loaded on to a 4–12% gradient Tris-Glycine gel (Invitrogen), separated via SDS–polyacrylamide gel electrophoresis (PAGE), and then transferred to HYBOND-P (PVDF) membrane (Amersham) at 30 V for 2 h. Membranes were rinsed and blocked overnight at 4°C . Membranes were probed with antibodies against GABA transporter 3 (rabbit anti-GAT-3 1:1,000; Millipore), and GABA transporter 1 (rabbit anti-GAT-1 1:200; Millipore). Following successive washes, membranes were incubated in IgG donkey anti-rabbit horseradish peroxidase (HRP)-labelled secondary (1:6,000; Jackson) for 1 h at room temperature ($21\text{--}23^{\circ}\text{C}$). Membranes were incubated in ECL PLUS (Amersham) and chemiluminescence was detected using Fluorochem (Alpha Innotech). Membranes were then re-probed for 1 h at room temperature ($21\text{--}23^{\circ}\text{C}$) with GAPDH (1:2,500; Abcam) and donkey anti-rabbit-HRP (1:10,000; Jackson) as an endogenous control protein to ensure equal loading. Immunoblotting was performed in triplicate for each antibody. Adobe Photoshop software (Adobe Systems) was used for densitometric analysis of all blots.

31. Lee, K., Kim, J. E., Sivula, M. & Strittmater, S. M. Nogo receptor antagonism promotes stroke recovery by enhancing axonal plasticity. *J. Neurosci.* **24**, 6209–6217 (2004).
32. Glykys, J., Mann, E. O. & Mody, I. Which GABA_A receptor subunits are necessary for tonic inhibition in the hippocampus? *J. Neurosci.* **28**, 1421–1426 (2008).
33. STAIR. Recommendations for standards regarding preclinical neuroprotective and restorative drug development. *Stroke*. **30**, 2752–2758 (1999).
34. Fisher, M., Feuerstein, G., Howells, D. W., Hurn, P. D., Kent, T. A., Savitz, S. I. & Lo, E. H. Update of the stroke therapy academic industry roundtable preclinical recommendations. *Stroke*. **40**, 2244–2250 (2009).
35. Stell, B. & Mody, I. Receptors with different affinities mediate phasic and tonic GABA_A conductances in hippocampal neurons. *J. Neurosci.* **22**, RC223 (2002).
36. Verheugen, J. A., Fricker, D. & Miles, R. Noninvasive measurements of the membrane potential and GABAergic action in hippocampal interneurons. *J. Neurosci.* **19**, 2546–2555 (1999).
37. Barlow, R. Cumulative frequency curves in population analysis. *Trends Pharmacol. Sci.* **11**, 404–406 (1990).
38. Baskin, Y. K., Dietrich, W. D. & Green, E. J. Two effective behavioral tasks for evaluating sensorimotor dysfunction following traumatic brain injury in mice. *J. Neurosci. Methods* **129**, 87–93 (2003).
39. Schallert, T., Fleming, S. M., Leasure, J. L., Tillerson, J. L. & Bland, S. T. CNS plasticity and assessment of forelimb sensorimotor outcome in unilateral rat models of stroke, cortical ablation, parkinsonism and spinal cord injury. *Neuropharmacology* **39**, 777–787 (2000).
40. Moore, C. S. *et al.* Increased X-linked inhibitor of apoptosis protein (XIAP) expression exacerbates experimental autoimmune encephalomyelitis (EAE). *J. Neuroimmunol.* **203**, 79–93 (2008).

Sequential faulting explains the asymmetry and extension discrepancy of conjugate margins

César R. Ranero^{1*} & Marta Pérez-Gussinyé^{2*}

During early extension, cold continental lithosphere thins and subsides, creating rift basins. If extension continues to final break-up, the split and greatly thinned plates subside deep below sea level to form a conjugate pair of rifted margins. Although basins and margins are ubiquitous structures, the deformation processes leading from moderately extended basins to highly stretched margins are unclear, as studies consistently report that crustal thinning is greater than extension caused by brittle faulting^{1–4}. This extension discrepancy might arise from differential stretching of brittle and ductile crustal layers², but that does not readily explain the typical asymmetric structure of conjugate margins^{5,6}—in cross-section, one margin displays gradual thinning accompanied by large faults, and the conjugate margin displays abrupt thinning but smaller-scale faulting⁵. Whole-crust detachments, active from early in the rifting, could in theory create both thinning and asymmetry¹, but are mechanically problematical. Furthermore, the extension discrepancy occurs at both conjugate margins, leading to the apparent contradiction that both seem to be upper plates to a detachment fault^{7,8}. Alternative models propose that much brittle extension is undetected because of seismic imaging limitations caused either by subseismic-resolution faulting⁹, invisible deformation along top-basement 100-km-scale detachments⁸ or the structural complexity of cross-cutting arrays of faults³. Here we use depth-migrated seismic images to accurately measure fault extension and compare it with crustal thinning. The observations are used to create a balanced kinematic model of rifting that resolves the extension discrepancy by producing both fault-controlled crustal thinning which progresses from a rift basin to the asymmetric structure, and extreme thinning of conjugate rifted margins. Contrary to current wisdom, the observations support the idea that thinning is to a first degree explained by simple Andersonian faulting that is unambiguously visible in seismic data.

Our model is based on data from the west Iberia margin (WIM) and its conjugate Newfoundland margin (NFM) that are archetypal for highly extended non-volcanic margins and exemplify rifting of initially cold lithosphere^{8,10} (Fig. 1). We analysed multibeam bathymetry¹¹, new pre-stack depth-migrated images of the WIM, and re-processed and pre-stack depth-migrated data from NFM. Bathymetry shows large fault-bounded north–south striking ridges (Fig. 1b) indicating that large-scale structures imaged on seismic line IAM11 are essentially two dimensional and, thus, representative of the rifting process (Fig. 2).

The transition from the ~20-km-thick continental crust of Galicia Bank to mantle exhumation involves seven major faults (Fig. 2). The fault blocks have relatively simple geometries with clear syntectonic strata overlying pre-rift (or earliest syn-rift) sediment, indicating that large fault blocks have not been dismembered by superimposed episodes of large-scale faulting as previously suggested³. The large faults f1–f6, with 4–6-km throws, have angular relationships to pre-rift strata which indicate that they formed at ~65–55° angles and rotated to the current ~42–28° (Fig. 2). Those dips differ by only a few degrees from true dips

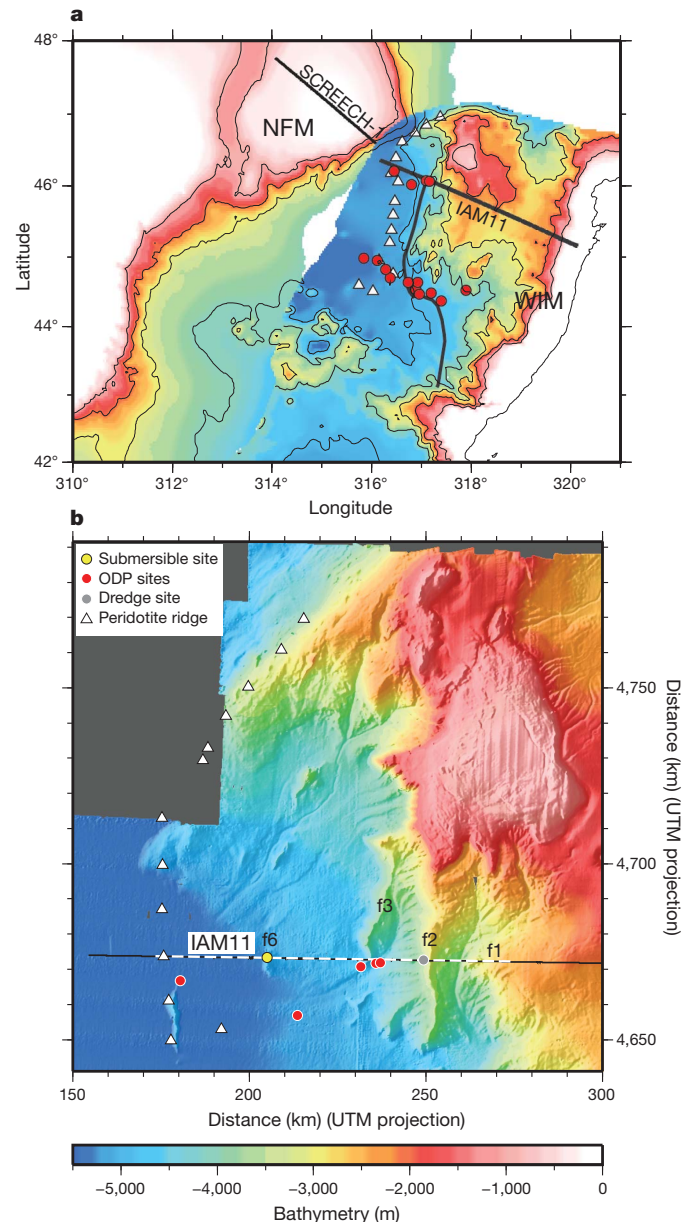


Figure 1 | WIM and NFM. **a**, Reconstruction at chron M029. Black lines are seismic profiles. IAM11 has been processed and pre-stack depth-migrated, and SCREECH-1 (originally published by ref. 5) has been reprocessed and pre-stack depth-migrated with refined velocities. The black curved line marks oceanward extent of continental crust at the WIM. **b**, Multibeam bathymetry of the deep Galicia margin¹¹. The IAM11 segment in Fig. 2 is marked by white dashes. Scarps f1, f2, f3 and f6 correspond to faults on IAM11 (Fig. 2a). ODP, Ocean Drilling Program. UTM, universal transverse Mercator.

¹ICREA at CSIC, Barcelona Center for Subsurface Imaging, Instituto de Ciencias del Mar, CSIC, Passeig Marítim de la Barceloneta 37-49, 08003, Barcelona, Spain. ²Department of Earth Sciences, Royal Holloway, University of London, Egham TW20 0BD, UK.

*These authors contributed equally to this work.

for f3, where the seismic profile is not exactly perpendicular to the fault plane (Fig. 1). The fault angles are compatible with Andersonian mechanics, which allows slip until an angle of $\sim 30^\circ$ is reached.

Bathymetry shows that WIM faults are arcuate in plan view, implying that the cross-sectional dimensions of imaged fault blocks are influenced by profile location (Fig. 1b). Despite this, the trend is for fault-block thickness generally to decrease from B1 to B7, from about 18 to 5 km. Studies of normal-fault populations in nature, and numerical and analogue models, establish that fault-block dimensions decrease with decreasing brittle-layer thickness^{12,13}. Furthermore, numerical models of cold-lithosphere rifting¹⁴ and pressure–temperature–time paths of lower crustal rocks^{15,16} indicate that the lower crust progressively becomes brittle as the crust thins. Thus, the decrease in dimensions from block B1 to block B7 indicates that at the time of fault activity, the brittle layer—which may have included most or all of the crust—must have decreased in thickness from fault f1 to f7 (Fig. 2), and included lower crustal rocks.

The age of faulting at the WIM is debated. Stratigraphic interpretation of 1980s seismic images indicated a rifting ~ 25 million years (Myr) long¹⁷. But reinterpretation of the images concluded that syn-rift sediment could not be objectively identified¹⁸, and faulting might have lasted < 5 Myr. Superior modern images have permitted us a better identification of pre- and syn-rift strata (Methods and Supplementary

Fig. 1). Sedimentation during f2 activity is older than sedimentation to f4, indicating that movement along faults was not coeval (Fig. 2). Diachronous-fault activity has also been interpreted in the southern segment of the WIM¹⁹.

The above observations may be explained if faults f1–f7 were formed successively and functioned sequentially through time, with each fault causing concomitant crustal thinning. For example, f3 formed on crust that had been previously cut and thinned by fault f2, and f3 thus bounds a smaller crustal block than f2. In this manner, successive f1–f7 fault activity produces gradually smaller B1–B7 crustal blocks (Fig. 2). Had faults f1–f7 functioned synchronously, they would have cut sections of crust of similar thicknesses and bound blocks of similar dimensions, which is not the case.

Comparison of structures on IAM11 with conjugate NFM shows that rifting has created a regional-scale asymmetric configuration. In NFM, Moho is steeper, crustal thinning abrupt, faulting comparatively subdued, and pre-rift and syn-rift sediment less differentiated (Fig. 3). However, coeval tectonic processes must have configured WIM and NFM architecture.

We have constructed a balanced kinematic model consistent with our observations of conjugate, highly extended margins, and with observations of the preceding evolution of faulting in less extended basins. Seismic images provide crustal structure, so our model does not

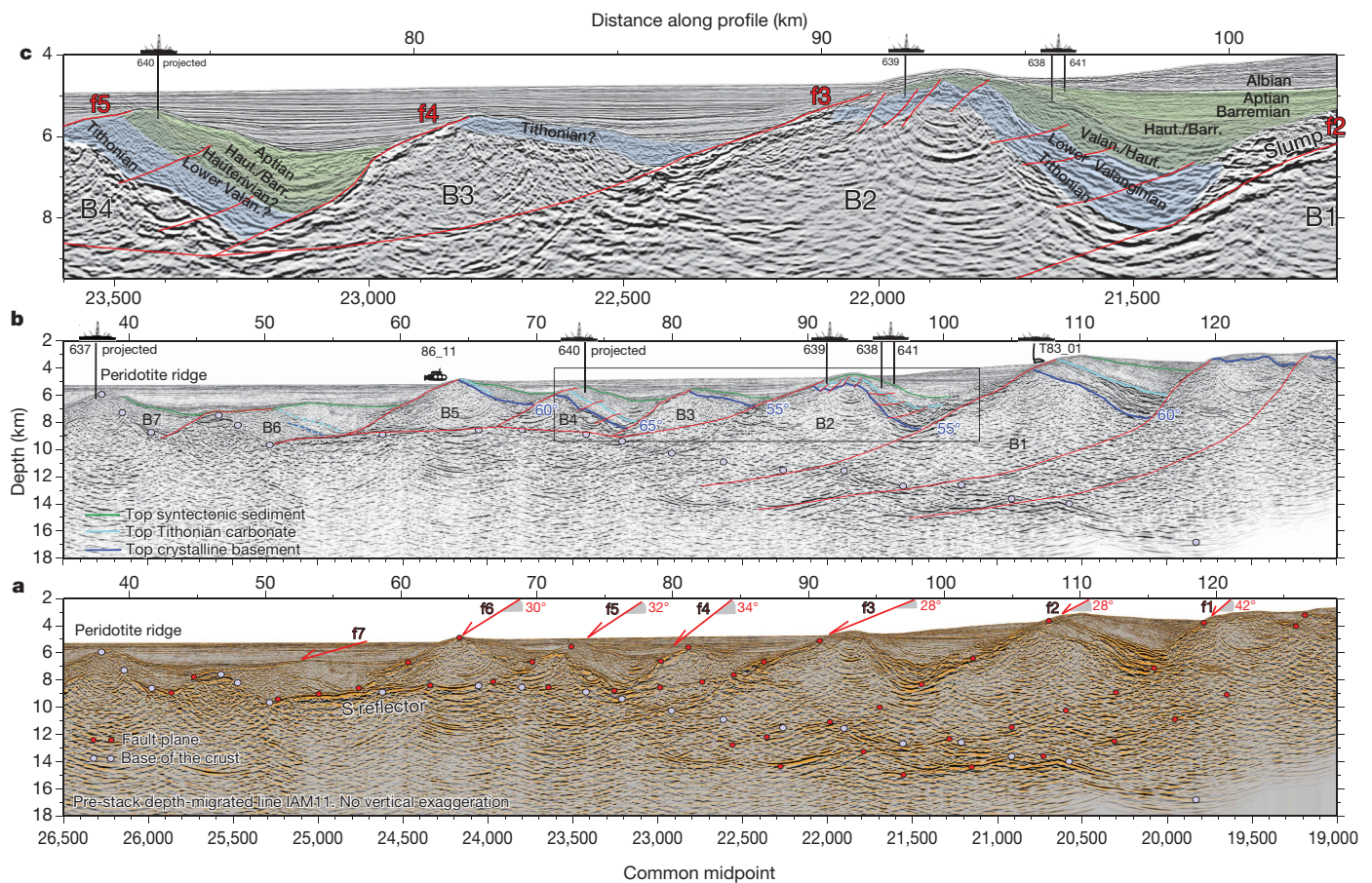


Figure 2 | Pre-stack depth-migrated line IAM11. **a**, Arrows and numbers indicate the average dips of the block-bounding fault segments exhumed during rifting. Red circles delineate reflections in crust and mantle interpreted as faults, which project up-dip to the exhumed block-bounding fault segments. Grey circles mark the crust base, constrained with wide-angle seismic data³⁰. **b**, Tectonic and stratigraphic interpretation calibrated with ODP sites 637–641, dive 86_11 and dredge T83_01 (ref. 17). Faults bounding blocks B1–B5 possibly formed at 55° – 65° (in blue), as inferred from their angular relationship to Jurassic strata originally deposited subhorizontally. **c**, Magnified view of box in **b**, showing geometry and calibrated ages of sediment units. At block B2, four

units overlie pre-rift carbonates. A Valanginian/Hauterivian unit, drilled at sites 638 and 639, was deposited during the main activity of fault f2. Younger Hauterivian/Barremian, Barremian/Aptian and Albian/recent units drilled at sites 638 and 641 are post-tectonic. Block B3 is imaged at its edge and sediment units are too thin to be resolved. At site 640, in block B4, Barremian/Aptian sediment was found as part of the section of syn-rift strata. Thus, the activity of fault f2 stopped during Hauterivian time, possibly before the activity of fault f4 began, during Barremian time, and several million years before the end of f4 activity, during late Aptian or early Albian time.

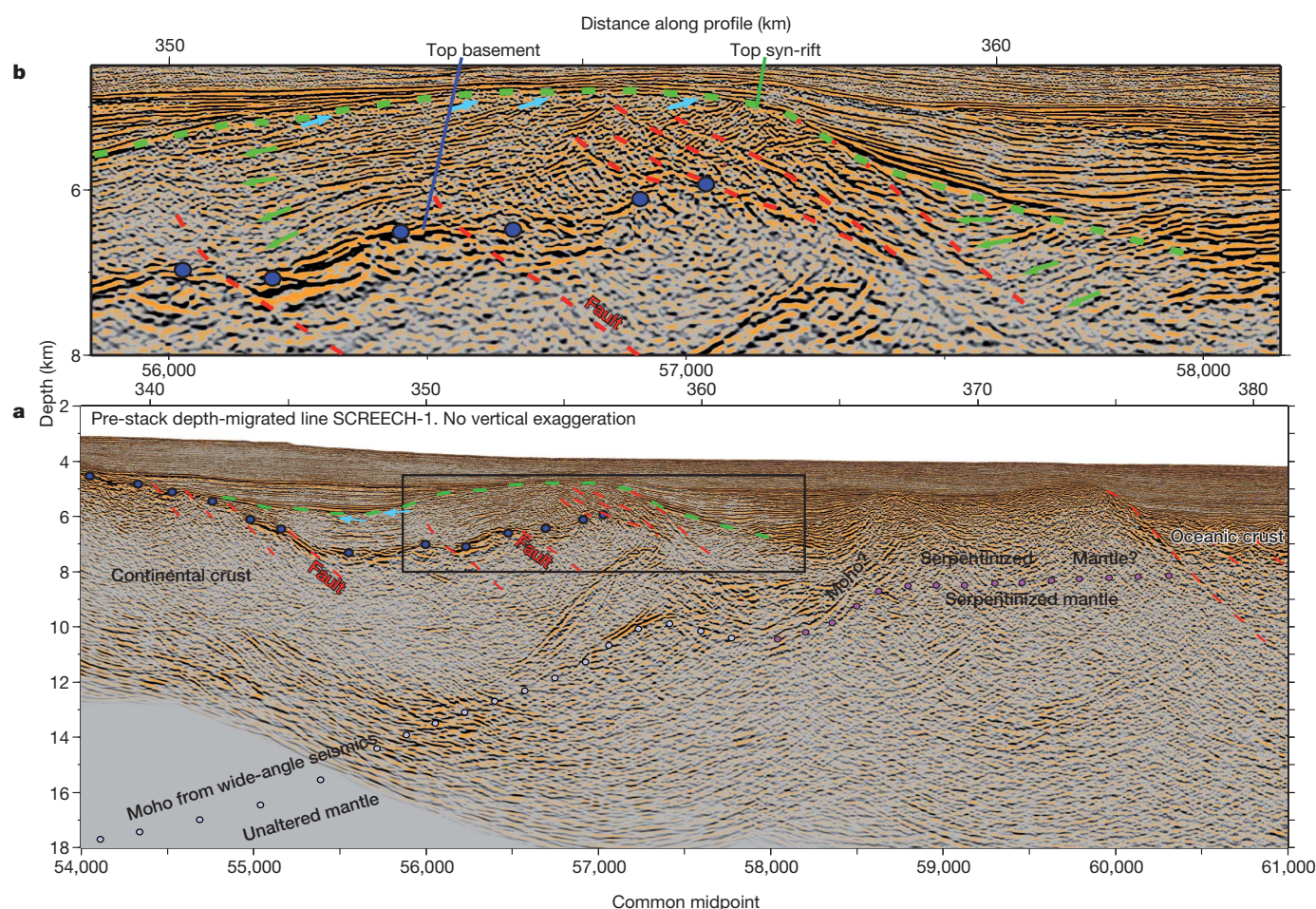


Figure 3 | Interpretation of a pre-stack depth-migrated segment of line SCREECH-1. a, The ~40-km-long segment illustrates the transition from continental to oceanic crust⁵, showing the extreme structural asymmetry

relative to the conjugated ~80-km-long section on IAM11 (Fig. 2).

b, Magnified view of the fault block marked by the box in **a**. Green arrows indicate strata fanning caused by syntectonic deposition.

consider lithospheric-mantle deformation. For basins, we use observations from regions where faulting is constrained in space and time, particularly the North Sea^{20,21} and the Gulf of Suez²², and from other well-studied regions such as the Gulf of Corinth²³. Our model differentiates between two stages, a rift basin stage (Fig. 4a–c) and a subsequent rifted margin stage (Fig. 4d–k), both of which consider the evolving rheology of extending crust. Model constraints are discussed in Methods and Supplementary Fig. 3 shows a stepwise construction.

Rift basin studies show that early extension is distributed across numerous disconnected, short faults with no dominant inward or outward sense of dip^{13,20}. In the WIM, this corresponds to shallow-water carbonate sedimentation found by drilling and dredging (Fig. 2). In this phase, faults rotate a few degrees (Fig. 4a). With increasing extension, faults grow laterally and link. Continued deformation focuses in a few active, larger faults. Simultaneously, outward-dipping faults halt, and strain concentrates in inward-dipping faults near the future basin centre, narrowing the area of active faulting^{20–22}. Faults in less thinned rift flanks become inactive (Fig. 4b). During this phase, the ductile lower crust¹⁴ deforms by localized thinning below the faulting region (Supplementary Fig. 2). Evidence for this comes from numerical models of cold lithosphere showing that the lower crust cannot flow long distances, but deforms in compliance with the upper crust^{24,25}. During the last phase of the rift basin stage, strain migrates to a single fault defining the basin centre²⁰. Whereas other faults stop at higher angles, a last fault (fc in Fig. 4c) works until it is no longer mechanically possible, and eventually all faults are locked.

However, stresses continue pulling plates apart and strain must be accommodated by new deformation. We propose this new phase as the

deformation style that characterizes the rifted margin stage. The hanging wall of the last active fault, fc, is sufficiently thin (~16.5 km; Fig. 4c) for most of the crust to behave brittly¹⁴, promoting strain localization²⁵. We propose that a new fault forms at the position favourable for strain localization (f1 in Fig. 4d), cutting the hanging wall to fc, slipping and further thinning the crust until it locks. Subsequent faults form sequentially, following the same pattern (Fig. 4e–k). To model the rifted margin stage, we use geometry, rotation angle and basement subsidence of faults f1–f7 measured on IAM11 (Fig. 2). After f1 stops, f2 forms in the hanging wall of f1. Slip and rotation on f2 deforms its footwall, changing the geometry of the deep segment of the inactive f1 from planar to slightly concave up (Fig. 4e–k), comparable to imaged fault geometry (Fig. 2a, b).

Sequential formation of single faults in the hanging wall of preceding faults, characterizing the rifted margin stage, efficiently thins the crust in comparison with an array of coeval faults, because successive single faults cut crust thinned beforehand by the preceding fault. It also leads to an asymmetric margin structure: one side of the model has inward-dipping faults and gentle Moho shoaling, resembling the WIM, and the conjugate side is characterized by more abrupt Moho shoaling and less faulting (Fig. 4k). To test whether the model reproduces the NFM structure, we include thinning by faulting at the NFM (Fig. 3). Although timing is not well constrained, assuming that faults worked sequentially yields a structure resembling the NFM (Fig. 4k). The model thus reproduces the structure of rifted margins, providing insight into several unresolved problems.

A first model outcome is that it resolves the discrepancy between horizontal extension factors attained by faulting (β_f) and crustal thinning factors^{2,3} (β_{cf}). If all faults function simultaneously, β_{cf} at highly

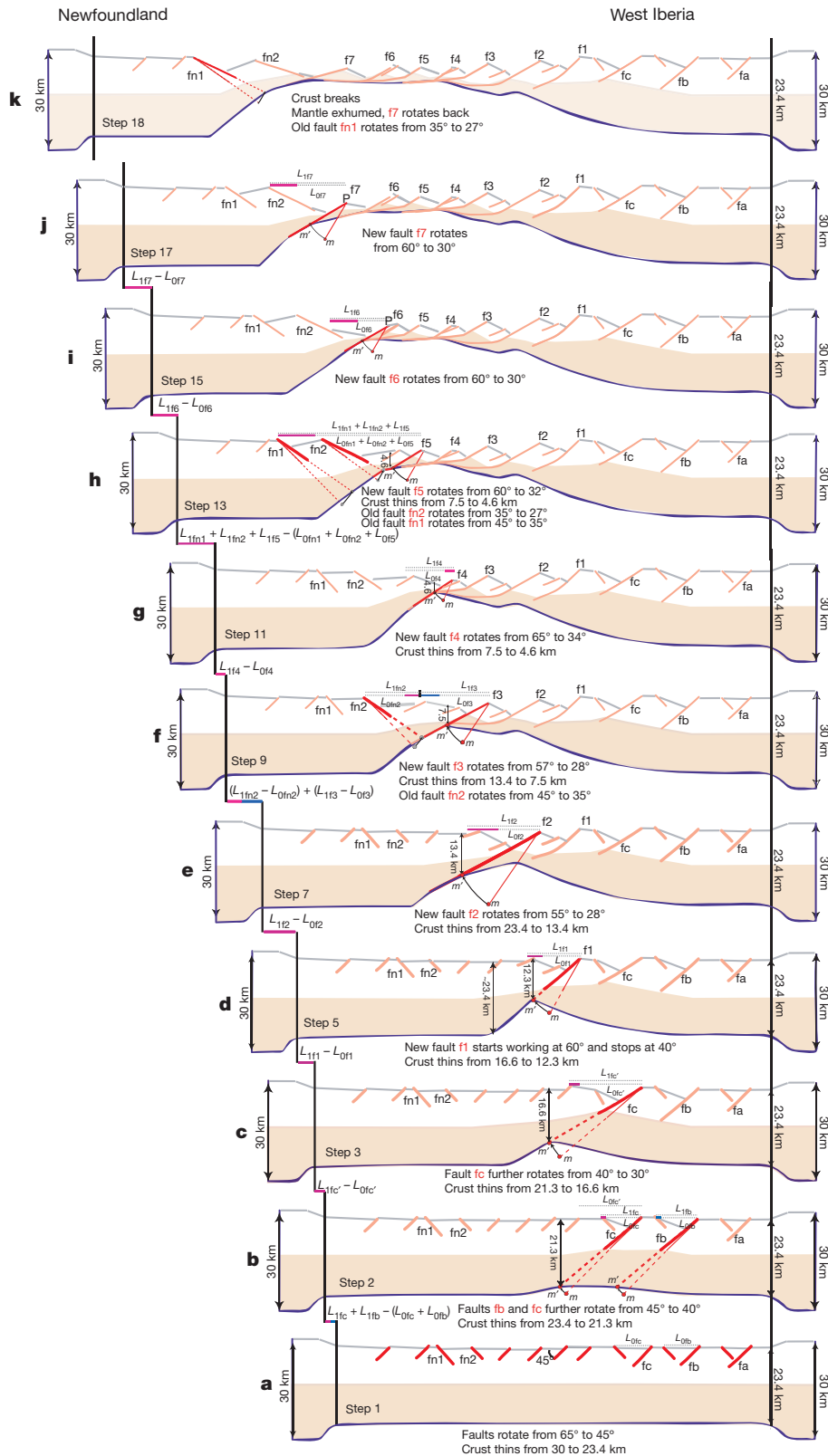


Figure 4 | Selected steps of the balanced cross-section of the kinematic model showing development of crustal thinning and margin asymmetry. Model is fixed at right-hand end. Basement is overlain by pre-rift sediment (in grey). Red and orange lines are active and inactive faults, respectively. Thin red lines mark the initial fault geometry. Black vertical arrows to the left of faults are crustal thicknesses. Blue line marks the Moho. Red dashed lines are projection of faults through ductile crust, used to geometrically estimate Moho movement (Supplementary Fig. 2). Fault activity and rotation angles in a–c are based on rift basins²⁰, and those in d–k were measured on IAM11 (Fig. 2). Pale orange shading

indicates petrological lower crust, which switches from ductile (a–c) to progressively more brittle at the basin centre¹⁴ (d–k). At each step, horizontal extension equals extension by active faults (indicated by pink and blue horizontal lines). L_{0fi} and L_{1fi} are the horizontal lengths of a fault block before and after extension of fault fi ($i = a, b, c, 1, 2, 3, 4, 5, 6, 7, n1, n2$). $L_{0fc'}$ and $L_{1fc'}$ are the horizontal lengths of the hanging-wall block to fault fc before and, respectively, after the fault movement occurring between steps 2 and 3. All reconstruction steps (including some omitted here) are shown in Supplementary Fig. 3. Geometrical rules are described in Supplementary Fig. 2 and Methods.

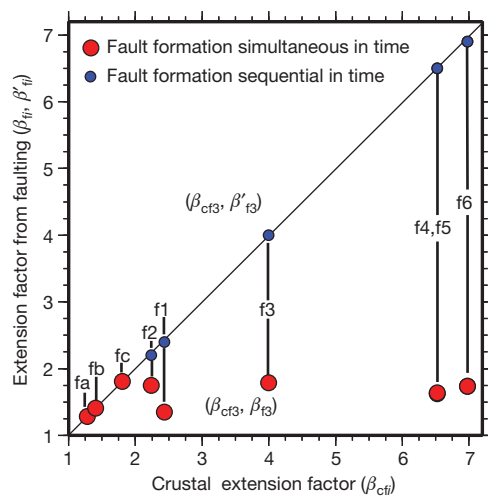


Figure 5 | Extension factors from model in Fig. 4. Crustal thinning factor (β_{cf}) versus horizontal extension factors from faulting (β_{fi} , red) and combined horizontal extension factor from sequential faulting (β'_{fi} , blue): β_{cf} is the ratio of initial thickness ($Z_0 = 30$ km) to crustal thickness in the hanging wall to fault f_i , $\beta_{fi} = L_{1fi}/L_{0fi}$ (Fig. 4) and $\beta_{cf} > \beta_{fi}$. If faults function simultaneously, β_{fi} cannot be reconciled with β_{cf} for faults f1–f6, giving rise to the so-called extension paradox. If faults function sequentially, as in Fig. 4d–k, each new fault cuts through crust that has already been thinned. To account for sequentially, we define the combined extension factor: $\beta'_{fi} = \prod_{j=n,i-1} \beta_{fj}$. Here $\beta_{f1}, \dots, \beta_{fi-1}$ are horizontal extension factors of faults that contributed to crustal thinning in the immediate footwall of f_i . In our model, crustal thinning is prescribed to equal horizontal extension at each time step (Supplementary Fig. 2), such that sequential faulting results in β'_{fi} equalling β_{cf} , mapping along the line of no extension discrepancy.

stretched margins (with $\beta_{cf} > 1.8$) is greater than the extension factor achievable with Andersonian faulting (Fig. 5). However, faults forming sequentially when $\beta_{cf} > 1.8$ resolve the paradox, because each fault cuts the hanging wall of the previous fault, adding localized further extension, such that the combined extension factor β'_{fi} is equal to β_{cf} (Fig. 5). Crustal thinning of NFM is not explained by faults imaged there, because it was caused by faults currently located in the WIM (Fig. 4d–k).

A second model outcome concerns the asymmetric structure characteristic of conjugate margins^{5,6}, which is typically explained by extension along long-lasting, crustal-scale detachment faults¹ active from early in the rifting at mechanically difficult low angles²⁶. In our model, asymmetry arises from strain localization promoting one direction of faulting during the rifted margin stage, as imaged on IAM11. We speculate that rheology controls the transition from distributed to localized extension. During early rifting, thick crust with ductile lower crust tends to distribute deformation on small faults over a wide area. Increasing thinning progressively makes the lower crust—and, thus, the entire crust—brittle¹⁴, promoting strain localization in the same way as at slow-spreading ocean ridges, where most extension occurs on a dominant fault cutting towards the thinnest lithosphere at the rift centre. Furthermore, the model predicts exhumation of lower crust to successively shallower levels by faults f1–f7. Evidence in support of this comes from rocks drilled at the edge of the continental crust in the WIM¹⁵.

A related outcome concerns the style of faulting during break-up. In the WIM, a subhorizontal structure imaged on profiles (for example underlying blocks B4–B7 in Fig. 2) has been interpreted as a detachment fault²⁷. Similar interpretations in other areas of the WIM and elsewhere consider those structures to be critical during the late rifting^{5,8,10,19,28}. Sequential faulting creates similar low-angle structures by passively rotating segments of inactive faults contained within the footwall block of active faults dipping at $>28^\circ$ (Fig. 4 and Methods). Thus, Andersonian faulting remains a viable alternative to widespread detachment faulting controlling break-up.

A last corollary is that the model may resolve two sediment-deposition paradoxes. It has been proposed that syn-rift sediment is surprisingly scarce in the WIM for the time elapsed between pre-rift deposition and initiation of seafloor spreading¹⁸. Although this controversial interpretation may partially arise from poor imaging (see, for example, Supplementary Fig. 1), sequential faulting implies that the syn-rift sediment on every tilted block is different in age, such that the total syn-rift stratigraphic log is thicker than if all faults work coeval. Also, the model naturally explains the observation that pre-rift or early syn-rift sediments deposited in shallow water, that is, on little-thinned crust, are found overlying extremely thin crust at great water depth^{17,24} (Fig. 4).

In contrast to conventional wisdom, a model based on structures unambiguously imaged by seismic methods explains the architecture of conjugate rifted margins formed by extension of cold lithosphere, showing that crustal thinning and asymmetry can both be explained to a first degree by Andersonian faulting.

METHODS SUMMARY

The greater penetration and higher vertical (temporal) and horizontal (spatial) resolution of modern multichannel seismic reflection data, the correct geometry of pre-stack depth-migrated images (Supplementary Fig. 1) and seismic velocity information obtained during depth migration have allowed us to re-evaluate the stratigraphy of fault blocks. The geometry of strata above the S reflector was recently studied²⁷, and we have extended the analysis to sediment units across the segment of the deep Galicia margin where most drilling occurred, and calibrated the images with Ocean Drilling Program Leg 103 results¹⁷ (Fig. 2c).

Additionally, we have developed a methodology for the kinematic reconstruction of crustal-scale balanced cross-sections that accounts for brittle and initial ductile deformation of the lower crust, subsequent lower-crust embrittlement during rifting¹⁴, progressive focusing of faulting towards the basin centre^{20,22} and the geometry of seismically imaged structures. Commercially available balancing software requires a fault detachment level, decoupling upper crustal deformation from lower crustal deformation, and cannot simulate the fault-controlled crustal-scale thinning imaged in our data. We achieve this crustal-scale thinning by moving Moho each extension step to a position where thinning across the active fault equals horizontal extension by the fault (Supplementary Fig. 2). This is done by rotating the point where the fault, or its projection, intersects the Moho by an angle equal to the fault rotation angle (Supplementary Fig. 2). The reconstruction (Supplementary Fig. 3) follows simple geometrical rules, described in Methods and Supplementary Fig. 2, producing a balanced cross-section with area conservation of better than 90%. The final model (Fig. 4k and Supplementary Fig. 3r) resembles the observed asymmetric architecture of conjugate margins (Supplementary Fig. 4). Our model underestimates thinning in NFM, possibly because seismic line SCREECH-1 is not perfectly conjugate to IAM11 (plate reconstruction uncertainties are tens of kilometres), and is somewhat oblique to faults, with the result that imaged fault dip is apparent and extension by faulting is underestimated.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 February; accepted 14 September 2010.

Published online 3 November 2010.

1. Lister, G. S., Etheridge, M. A. & Symonds, P. A. Detachment models for the formation of passive continental margins. *Tectonics* **10**, 1038–1064 (1991).
2. Davis, M. & Kusznir, N. J. in *Proc. NSF Rifted Margins Theor. Inst.* (ed. Karner, G. D.) 92–136 (Columbia Univ. Press, 2004).
3. Reston, T. J. Extension discrepancy at North Atlantic nonvolcanic rifted margins: depth-dependent stretching or unrecognized faulting? *Geology* **35**, 367–370 (2007).
4. Sibuet, J. C. New constraints on the formation of the non-volcanic continental Galicia–Flemish Cap conjugate margins. *J. Geol. Soc. Lond.* **149**, 829–840 (1992).
5. Hopper, J. R. *et al.* Continental break-up and the onset of ultraslow seafloor spreading off Flemish Cap on the Newfoundland rifted margin. *Geology* **32**, 93–96 (2004).
6. Loudon, K. E. & Chian, D. The deep structure of non-volcanic rifted continental margins. *Phil. Trans. R. Soc. Lond. A* **357**, 767–804 (1999).
7. Driscoll, N. W. & Karner, G. D. Lower crustal extension across the Northern Carnarvon basin, Australia: evidence for an eastward dipping detachment. *J. Geophys. Res.* **103**, 4975–4991 (1998).
8. Lavier, L. & Manatschal, G. A mechanism to thin the continental lithosphere at magma-poor margins. *Nature* **440**, 324–328 (2006).
9. Marrett, R. & Allmendinger, R. Amount of extension on “small” faults: an example from the Viking graben. *Geology* **20**, 47–50 (1992).

10. Whitmarsh, R. B., Manatschal, G. & Minshull, T. A. Evolution of magma-poor continental margins from rifting to seafloor spreading. *Nature* **413**, 150–154 (2001).
11. Ercilla, G. *et al.* Contourites in the Galicia Bank region (NW Iberian Atlantic). *Mar. Geophys. Res.* (in the press).
12. Soliva, R., Benedicto, A. & Maerten, L. Spacing and linkage of confined normal faults: importance of mechanical thickness. *J. Geophys. Res.* **111**, B01402 (2006).
13. Ackermann, R. V., Schlische, R. W. & Withjack, M. O. The geometric and statistical evolution of normal fault systems: an experimental study of the effects of mechanical layer thickness on scaling laws. *J. Struct. Geol.* **23**, 1803–1819 (2001).
14. Pérez-Gussinyé, M. & Reston, T. J. Rheological evolution during extension at passive non-volcanic margins: onset of serpentinization and development of detachments leading to continental break-up. *J. Geophys. Res.* **106**, 3961–3975 (2001).
15. Gardien, V. *et al.* The evolution of amphibolites from Site 1067, ODP Leg 173 (Iberia Abyssal Plain): Jurassic rifting to the Pyrenean compression. *Spec. Publ. Geol. Soc. (Lond.)* **187**, 191–208 (2001).
16. Muentener, O., Hermann, J. & Tromsdoerff, V. Cooling history and exhumation of lower-crustal granulite and upper mantle (Malenco, eastern central Alps). *J. Petrol.* **41**, 175–200 (2001).
17. Boillot, G. & Winterer, E. L. Drilling on the Galicia Margin: retrospect and prospect. *Proc. ODP Sci. Results* **103**, 809–828 (1988).
18. Wilson, R. C. L., Manatschal, G. & Wise, S. Rifting along non-volcanic passive margins: stratigraphic and seismic evidence from the Mesozoic of the Alps and Western Iberia. *Spec. Publ. Geol. Soc. (Lond.)* **187**, 429–452, 2001.
19. Péron-Pinvidic, G., Manatschal, G., Minshull, T. A. & Sawyer, D. S. Tectonosedimentary evolution of the deep Iberia-Newfoundland margins: evidence for a complex break-up history. *Tectonics* **26**, TC2011 (2007).
20. Cowie, P. A., Underhill, J. R., Behn, M. D., Lin, J. & Gill, C. E. Spatio-temporal evolution of strain accumulation derived from multi-scale observations of Late Jurassic rifting in the northern North Sea: a critical test of models for lithospheric extension. *Earth Planet. Sci. Lett.* **234**, 401–419 (2005).
21. Walsh, J. J. *et al.* Strain localisation and population changes during fault system growth within the Inner Moray Firth, northern North Sea. *J. Struct. Geol.* **25**, 307–315 (2003).
22. Gawthorpe, R. L. *et al.* Normal fault growth, displacement localisation and the evolution of normal fault populations: the Hamman Faraun fault block, Suez Rift, Egypt. *J. Struct. Geol.* **25**, 1347–1348 (2003).
23. Mattei, M. *et al.* Tectonic evolution of fault-bounded continental blocks: comparison of paleomagnetic and GPS data in the Corinth and Megara basins (Greece). *J. Geophys. Res.* **109**, B02106 (2004).
24. Hopper, J. & Buck, R. Styles of extensional decoupling. *Geology* **26**, 699–702 (1998).
25. Huismans, R. S. & Beaumont, C. Symmetric and asymmetric lithospheric extension: Relative effects of frictional-plastic and viscous strain softening. *J. Geophys. Res.* **108**, 2496 (2003).
26. Sibson, R. H. Fluid involvement in normal faulting. *J. Geodyn.* **29**, 469–499 (2000).
27. Reston, T. J. *et al.* Movement along a low-angle normal fault: the S reflector west of Spain. *Geochem. Geophys. Geosyst.* **8**, Q06002 (2007).
28. Manatschal, G. New models for evolution of magma-poor rifted margins based on a review of data and concepts from West Iberia and the Alps. *Int. J. Earth Sci.* **93**, 432–466 (2004).
29. Srivastava, S. P., Sibuet, J.-C., Cande, S., Roest, W. R. & Reid, I. R. Magnetic evidence for slow spreading during the formation of the Newfoundland and Iberian margins. *Earth Planet. Sci. Lett.* **182**, 61–76 (2000).
30. Whitmarsh, R. B. *et al.* The ocean-continent boundary off the western continental margin of Iberia: crustal structure west of Galicia Bank. *J. Geophys. Res.* **101**, 28291–28314 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The ideas presented in this work have benefited from discussions with J. Adam, A. Amilibia, E. Casciello, T. Cunha, M. R. Fowler, S. Hardy, J. García-Pintado, G. Manatschal, K. McClay, M. Menzies, E. Saura and F. Storti, and from numerous discussions and previous collaborations with T. J. Reston. The early work that led to the ideas presented here was carried out when C.R.R. worked at IFM-GEOMAR and M.P.-G. at IFM-GEOMAR and later at ICTJA-CSIC. We are grateful to J. Collier and C. Beaumont for their reviews, which helped improved this article. This is a publication of the Department of Earth Sciences of the Royal Holloway, University of London. C.R.R. has been supported by the Kaleidoscope project, funded by Repsol, and by the Spanish National Project Medoc of the Ministry of Science and Innovation.

Author Contributions C.R.R. processed the seismic data up to pre-stack depth migration and interpreted them. M.P.-G. built the tectonic model. Both authors contributed equally to writing the manuscript and to developing the ideas behind the tectonic model.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.R.R. (cranero@icm.csic.es) or M.P.-G. (m.perez@es.rhul.ac.uk).

METHODS

Pre-stack depth-migrated seismic reflection records have allowed the re-evaluation of both the fault-block sediment stratigraphy in the segment of the deep Galicia margin where most drilling has occurred¹⁷ and the tectonic structure formed during rifting. On the basis of this reinterpretation, we have developed a methodology for the kinematic reconstruction of crustal-scale balanced cross-sections.

Kinematic reconstruction. A first set of reconstruction rules (rules 1–6) was imposed to define the evolution of deformation in the crustal scale. These rules are based on published, reasonably well-accepted constraints (rules 1–5), and on the geometry of structures imaged on seismic lines IAM11 and SCREECH-1 (rule 6). We describe these rules and how they define how extension progresses to final continental break-up.

(1) At the start of rifting, the crust is ~30 km thick and has a temperature of ~500 °C at the Moho, which is appropriate for rifting conditions at the Galicia margin^{8,14,16}. Numerical modelling following ref. 14 shows that given those initial conditions and an appropriate extension velocity for the west Iberia/Newfoundland system, the lower crust (orange shading in Fig. 4, and Supplementary Figs 2–4) progressively becomes brittle, and becomes entirely brittle when the total crust is ~12–14 km thick (Supplementary Fig. 3h–i). (2) During early rifting phases, at the basin stage (Fig. 4a–c and Supplementary Fig. 3a–c), strain in faults gradually focuses into the basin centre^{20,22}. (3) During this basin stage, faults start at 65° and rotate to 45°–30° following Andersonian principles^{31,32} until all faults lock up (Fig. 4a–c and Supplementary Fig. 3a–c). (4) We assume that the following margin stage (Fig. 4d–k and Supplementary Fig. 3d–r) starts with the creation of a new fault, after all previous faults have locked up, located where the crust is thinnest (Supplementary Fig. 3d). We choose that location because deformation is expected to focus on the segment with the smallest elastic thickness. (5) During the progression of extension, as the lower crust switches from ductile to brittle deformation, it is first partially and later entirely cut by faults (faults f1–f7 in Fig. 4d–k and Supplementary Fig. 3d–r), and is exhumed at the thinnest fault blocks (faults f5–f7 in Fig. 4h–k and Supplementary Fig. 3m–r) as observed at drilled continental margins and exhumed margin analogues²⁸. (6) To create the margin stage geometry, fault spacing, initial and final fault rotation angles and basement subsidence are all directly taken from measured structures imaged at profile IAM11 (Fig. 2). In this manner, isostasy of large-offset faults is embedded in the reconstruction.

The second set of reconstruction rules (rules 7–15) is imposed to maintain common geometric rules in all steps of the model, and conforms with the deformation principles defined by the first set of rules.

(7) Large active faults are planar in the brittle domain^{33,34}. (8) During model construction, we choose horizontal extension to be accomplished by rotation of the active faults. Each fault rotates with respect to an axis at its top end. It can be proved that the resulting geometry of the crust and fault is independent of the location chosen for the rotation axis along the fault. We choose the fault top as the rotation axis to simplify the reconstruction. (9) The horizontal extension of the model at a time step is equal to the cumulative horizontal extension of all faults active at that time step. (10) A fault has the same length along the hanging wall and the footwall. (11) During faulting, the top and base of brittle hanging-wall crust are offset by faulting, but when the lower crust is ductile, fault offset attenuates across the ~3-km-thick brittle ductile transition (BDT; Supplementary Fig. 2). (12) The lower crust undergoes deformation beneath the segments where the upper crust extends, such that lower crustal deformation changes from a broad region (Fig. 4a, b and Supplementary Fig. 3a, b) to a narrow region in parallel with the change in strain distribution from multiple faults to one fault as extension progresses (Fig. 4c–k and Supplementary Fig. 3c–r). (13) Crustal thinning by an active fault equals horizontal extension by the fault. This is attained by rotating the location where (the projection of) the fault intersects the Moho by an angle equal to the rotation angle of the fault (Supplementary Fig. 2). (14) We assume that block deformation decreases away from faults. Footwall deformation is approximated by an affine transformation involving shear and stretch of triangle POM' (Supplementary Fig. 2), which is a reasonable geometrical approximation to real footwall deformation³⁵. The distance over which the footwall deformation associated with fault slip decreases to zero possibly depends on the elastic thickness and, thus, on the evolving rheology. However, there is no established geometric principle to determine how far footwall deformation extends, and we considered two modes depending on crustal rheology. When the lower crust is ductile, footwall deformation occurs underneath the fault plane and the area increase from POM to POM' is accomplished by ductile deformation (Supplementary Figs 2a–c and 3a–d). When the entire crust is brittle (Supplementary Figs 2d, e and 3f, r), footwall deformation affects the active fault footwall block and extends to older fault blocks located underneath the active fault. Here the area increase from POM

to POM' is <3%, a result easily explained by subseismic internal block deformation⁹. Therefore, changes in fault geometry from planar to listric occurring in the thinner crust of the margin stage result from block deformation (see, for example, Supplementary Figs 2d, e and 3f–r). (15) The ductile parts of the lower crust in the hanging wall deform to account for extension with area preservation (see Supplementary Fig. 2 for details) in a manner similar to the cantilever model³⁶; but in that model all faults rotate simultaneously, so it does not explain the extension discrepancy at high stretching factors.

The reconstruction reproduces the asymmetric geometry imaged on IAM11 and SCREECH-1 (Supplementary Fig. 4), and the regional asymmetry observed at other non-volcanic (cold-lithosphere) conjugate margins surveyed using coincident deep-penetrating wide-angle seismic data and multichannel seismic reflection images, such as the Porcupine basin rift flanks³⁷, south Labrador Sea/Greenland^{38,39}, Armorica/Flemish Cap^{6,38}, the Galicia Interior basin flanks⁴⁰ and Nova Scotia/Morocco⁴¹, and even magmatically robust systems such as the Gulf of Aden⁴² and Woodlark basin⁴³, are clearly asymmetrical.

Fault and coeval Moho rotation (at, for example, f5; Supplementary Fig. 3m) may cause rotation of the deep segment of a neighbouring inactive fault (for example f4 in Supplementary Fig. 3l). In thick crust, large faults tend to be distant, so fault rotation does not significantly affect the planar geometry of older inactive ones (fa, fb and fc; Fig. 4a–c and Supplementary Fig. 3a–c). As the crust thins and faults form at smaller spacing, rotation of an active fault modifies the geometry of the plane of a neighbouring inactive fault in its deep segment. The inactive fault plane is passively rotated to a smaller angle, producing an overall listric fault geometry, or even a subhorizontal appearance (Fig. 4e–k and Supplementary Fig. 3g–r). In particular, our model reproduces the break-away and subhorizontal geometry of a previously interpreted detachment fault. But our model does not require a mechanically controversial very-low-angle fault activity (the S reflector, imaged under blocks B4–B7 in Fig. 2, corresponding to blocks bounded by faults f4–f6 in Fig. 3 and Supplementary Fig. 3).

The kinematics of fault blocks B4–B7 might resemble the 'roll and hinge' model⁴⁴. However, our model is different because it has no hinge fixed in space from which successive slivers of crust are sliced. Additionally, the model reproduces—using one common Andersonian mechanistic framework—the change in fault geometry that has been related to increasing crustal extension at margins: the model evolves from deformation at distributed planar faults in little-extended sectors to deformation on faults that, owing to passive rotation, rest at low angles at highly extended sectors.

- Anderson, E. M. The dynamics of faulting. *Trans. Edinb. Geol. Soc.* **8**, 387–402 (1905).
- Célérier, B. Seeking Anderson's faulting in seismicity: a centennial celebration. *Rev. Geophys.* **46**, RG4001 (2008).
- Thatcher, W. & Hill, D. P. A simple model for the fault-generated morphology of slow-spreading mid-ocean ridges. *J. Geophys. Res.* **100**, 561–570 (1995).
- Jackson, J. A. Active normal faulting and crustal extension. *Spec. Publ. Geol. Soc. (Lond.)* **28**, 3–17 (1987).
- Westaway, R. & Kusznir, N. Fault and bed 'rotation' during continental extension: block rotation or vertical shear? *J. Struct. Geol.* **15**, 753–770 (1993).
- Kusznir, N. J., Marsden, G. & Egan, S. S. A flexural-cantilever simple-shear/pure-shear model of continental lithosphere extension: applications to the Jeanne d'Arc Basin, Grand Banks and Viking Graben, North Sea. *Spec. Publ. Geological Society (Lond.)* **56**, 41–60 (1991).
- Reston, T. J. *et al.* Extreme crustal thinning in the south Porcupine Basin and the nature of the Porcupine Median High: implications for the formation of non-volcanic rifted margins. *J. Geol. Soc. Lond.* **161**, 783–798 (2004).
- Keen, C., Peddy, C., de Voogd, B. & Matthews, D. Conjugate margins of Canada and Europe: results from deep reflection profiling. *Geology* **17**, 173–176 (1989).
- Chian, D., Keen, C., Reid, I. & Loudon, K. E. Evolution of nonvolcanic rifted margins: new results from the conjugate margins of the Labrador Sea. *Geology* **23**, 589–592 (1995).
- Pérez-Gussinyé, M., Ranero, C. R., Reston, T. J. & Sawyer, D. Mechanisms of extension at nonvolcanic margins: evidence from the Galicia interior basin, west of Iberia. *J. Geophys. Res.* **108**, 2245 (2003).
- Maillard, A., Malod, J., Thiebot, E., Klingelhoefer, F. & Rehault, J.-P. Imaging a lithospheric detachment at the continent–ocean crustal transition off Morocco. *Earth Planet. Sci. Lett.* **241**, 686–698 (2006).
- Fournier, M., Huchon, P., Khanbary, K. & Leroy, S. Segmentation and along-strike asymmetry of the passive margin in Socotra, eastern Gulf of Aden: are they controlled by detachment faults? *Geochim. Geophys. Geosyst.* **8**, Q03007 (2007).
- Kington, J. D. & Goodliffe, A. M. Plate motions and continental extension at the rifting to spreading transition in Woodlark Basin, Papua New Guinea: can oceanic plate kinematics be extended into continental rifts? *Tectonophysics* **458**, 82–95 (2008).
- Buck, W. R. Flexural rotation of normal faults. *Tectonics* **7**, 959–973 (1988).

Integrating carbon–halogen bond formation into medicinal plant metabolism

Weerawat Runguphan^{1*}, Xudong Qu^{1†*} & Sarah E. O'Connor¹

Halogenation, which was once considered a rare occurrence in nature, has now been observed in many natural product biosynthetic pathways¹. However, only a small fraction of halogenated compounds have been isolated from terrestrial plants². Given the impact that halogenation can have on the biological activity of natural products¹, we reasoned that the introduction of halides into medicinal plant metabolism would provide the opportunity to rationally bioengineer a broad variety of novel plant products with altered, and perhaps improved, pharmacological properties. Here we report that chlorination biosynthetic machinery from soil bacteria can be successfully introduced into the medicinal plant *Catharanthus roseus* (Madagascar periwinkle). These prokaryotic halogenases function within the context of the plant cell to generate chlorinated tryptophan, which is then shuttled into monoterpene indole alkaloid metabolism to yield chlorinated alkaloids. A new functional group—a halide—is thereby introduced into the complex metabolism of

C. roseus, and is incorporated in a predictable and regioselective manner onto the plant alkaloid products. Medicinal plants, despite their genetic and developmental complexity, therefore seem to be a viable platform for synthetic biology efforts.

Numerous halogenase enzymes from soil bacteria have been identified and characterized extensively^{1,3–5}. Two of these flavoenzymes, PyrH^{6,7} and RebH^{8–11}, chlorinate the indole ring of tryptophan in the five and seven positions, respectively. Transferring these enzymes into other natural product pathways would allow site-specific incorporation of halogens onto a range of tryptophan-derived alkaloid products¹², provided that the downstream enzymes could accommodate the chlorinated tryptophan precursor.

Catharanthus roseus produces a wide variety of monoterpene indole alkaloids¹³ (Fig. 1a). This metabolic pathway begins with the conversion of tryptophan (1) to tryptamine (2) by tryptophan decarboxylase¹⁴. Tryptamine then condenses with the iridoid terpene

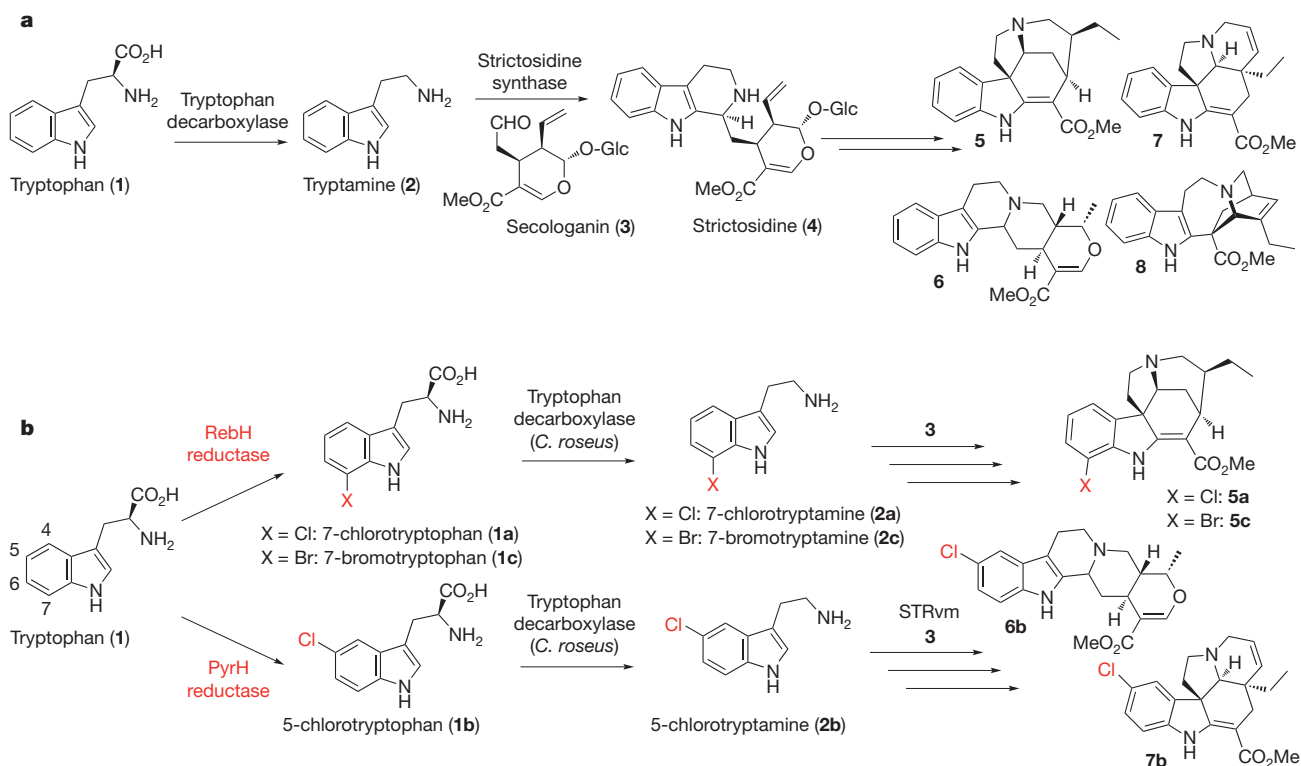


Figure 1 | Monoterpene indole alkaloid biosynthesis. **a**, Tryptophan (1) is decarboxylated by tryptophan decarboxylase to yield tryptamine (2), which reacts with secologanin (3) to form strictosidine (4). After numerous rearrangements, strictosidine (4) is converted into a variety of monoterpene indole alkaloids, such as 19,20-dihydroakuammicine (5), ajmalicine (6), tabersonine (7) and catharanthine (8). These compounds have a variety of

pharmacological activities^{24–26,30}. Me, CH₃; Glc, glucose. **b**, RebH and PyrH, along with a partner reductase, halogenate the indole ring of tryptophan to yield chlorotryptophan. Here we show that after transformation of these enzymes into *C. roseus*, the halogenated tryptophans **1a** and **1b** can be decarboxylated by tryptophan decarboxylase (*C. roseus*) to form the chlorotryptamines **2a** and **2b** and then converted into chlorinated monoterpene indole alkaloids.

¹Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [†]Present address: State Key Laboratory of Bioorganic and Natural Products Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, 345 Lingling Road, Shanghai 200032, China.

*These authors contributed equally to this work.

secologanin (**3**) to form a biosynthetic intermediate strictosidine (**4**), which is subsequently functionalized in *C. roseus* to form over 100 alkaloids, including the anticancer agent vinblastine¹³. Previous work has shown that when *C. roseus* cell culture is supplemented with a variety of halogenated tryptamines, the corresponding halogenated alkaloid analogues are produced in isolable yields^{15,16}. If prokaryotic halogenases could function in the eukaryotic plant cell, and if tryptophan decarboxylase could convert halogenated tryptophan into halogenated tryptamine, then *C. roseus* would produce chlorinated alkaloids *de novo* (Fig. 1b).

Because RebH and PyrH do not turn over tryptamine, this strategy requires that tryptophan decarboxylase from *C. roseus* recognize halogenated tryptophan. We assayed tryptophan decarboxylase from *C. roseus* *in vitro* with tryptophan ($K_m = 51.7 \pm 9.2 \mu\text{M}$ (Michaelis constant), $k_{\text{cat}} = 5.1 \pm 0.1 \text{ min}^{-1}$ (turnover number), $k_{\text{cat}}/K_m =$

$0.099 \mu\text{M}^{-1} \text{ min}^{-1}$), 7-chlorotryptophan (**1a**; $K_m = 499 \pm 74 \mu\text{M}$, $k_{\text{cat}} = 1.6 \pm 0.04 \text{ min}^{-1}$, $k_{\text{cat}}/K_m = 0.00327 \mu\text{M}^{-1} \text{ min}^{-1}$) and 5-chlorotryptophan (**1b**; $K_m = 538 \pm 48 \mu\text{M}$, $k_{\text{cat}} = 2.5 \pm 0.08 \text{ min}^{-1}$, $k_{\text{cat}}/K_m = 0.00455 \mu\text{M}^{-1} \text{ min}^{-1}$) (Supplementary Figs 1 and 2). The activity of the enzyme suggested that halogenated tryptophan could be decarboxylated *in vivo*.

When considering how to merge the prokaryotic biosynthetic machinery with the plant alkaloid pathway, we chose to transfer the halogenase enzymes into *C. roseus* rather than move the plant biosynthetic enzymes into a microbial host. Most of the monoterpene indole alkaloid biosynthetic genes have not been identified, making heterologous expression of this pathway impossible at present. Moreover, we note that reconstitution of plant alkaloid pathways continues to be a challenge^{17,18}. Many alkaloids use complex starting materials (such as secologanin) that are only produced by a few specialized plants, so

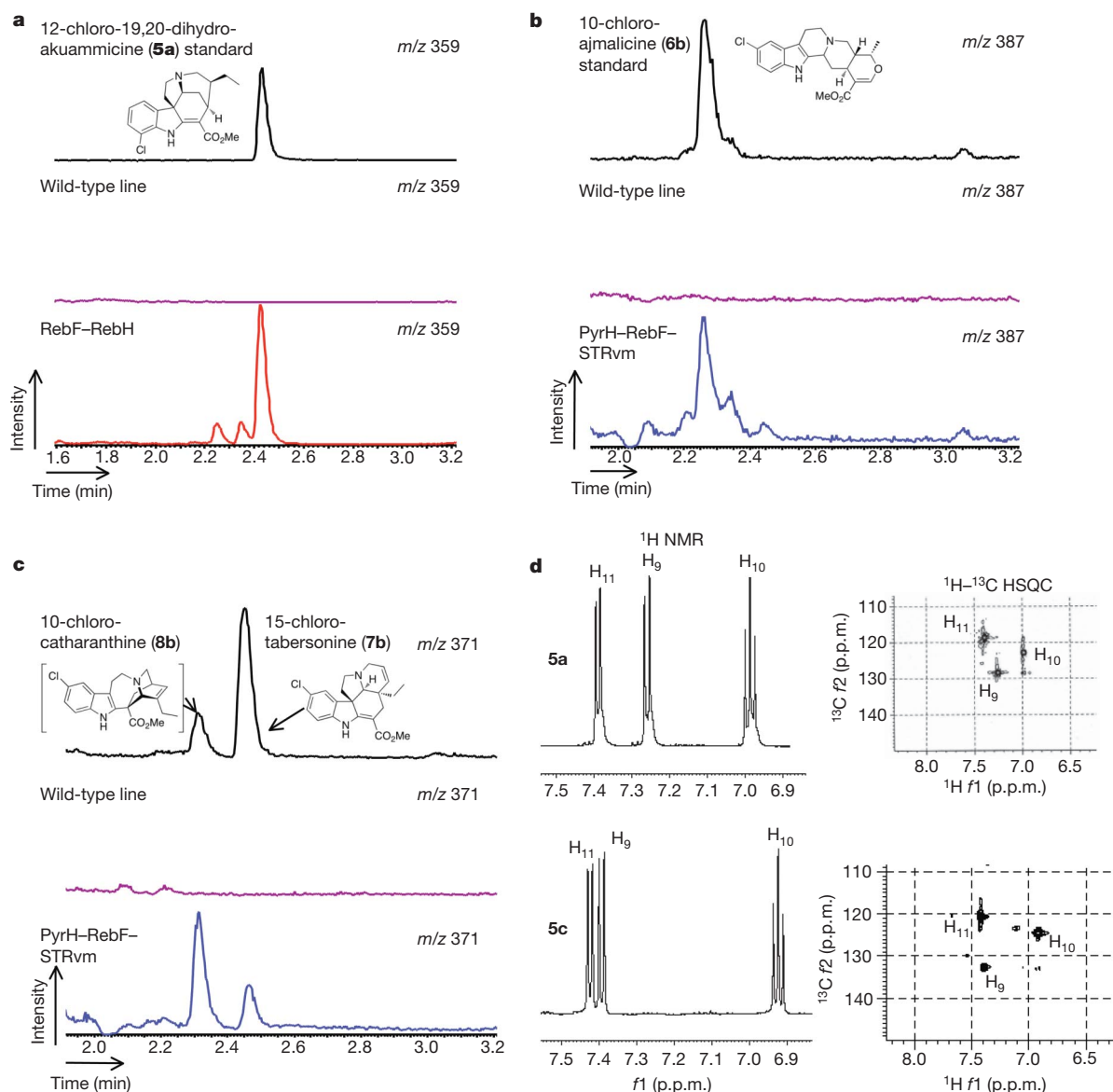


Figure 2 | Chlorinated alkaloids in *C. roseus* hairy root culture. **a**, LC–MS chromatograms showing 12-chloro-19,20-dihydroakuammicine (**5a**; m/z 359) in RebF–RebH hairy roots (red trace), contrasted with control cultures transformed with no plasmid (purple trace). An authentic standard of **5a** validated the structural assignment (black trace; Supplementary Figs 30 and 31). **b**, Chromatograms showing 10-chloroajmalicine (**6b**) in RebF–PyrH–STRvm hairy roots (blue trace), contrasted with control cultures (purple trace). An authentic standard of **6b** is shown²² (black trace). **c**, Chromatograms

showing 15-chlorotabersonine (**7b**) in RebF–PyrH–STRvm hairy roots (blue trace), contrasted with control cultures (purple trace). The other major peak at m/z 371 had an exact mass and ultraviolet spectrum consistent with a chlorinated analogue of catharanthine²² (**8**) (Supplementary Fig. 29). **d**, ^1H NMR and ^1H – ^{13}C heteronuclear single quantum coherence (HSQC) spectra of **5a** and **5c**. f_1 and f_2 , chemical shifts in the ^1H and ^{13}C dimensions, respectively.

reconstitution of plant alkaloid pathways must also include biosynthesis of these precursors. For example, ajmalicine (**6**; Fig. 1a), one of the simplest of the monoterpene indole alkaloids, requires an estimated 14 discrete enzymes for biosynthesis from tryptophan and the terpene geraniol¹³; reconstitution of a pathway of this length is a significant engineering problem. Therefore, we believe that exploring approaches in the host plant is an important aspect of alkaloid metabolic engineering efforts.

To produce 7-chlorotryptophan *in planta*, we generated an expression construct containing codon-optimized complementary DNA encoding the 7-tryptophan chlorinase RebH and its required partner flavin reductase, RebF, in a plant expression vector (pCambia1300), both under the control of constitutive cauliflower mosaic virus (CaMV) 35S promoters. For production of 5-chlorotryptophan, an expression construct encoding the 5-chlorinating enzyme PyrH, along with RebF as the partner reductase, was generated. No signal sequence was added to the halogenase genes, to ensure that RebH, PyrH and RebF would produce chlorinated tryptophan in the cytosol, where it would most readily encounter the decarboxylase, which is also localized in the cytosol¹⁹ (Supplementary Figs 3–5).

We used *Agrobacterium rhizogenes* to generate hairy root culture of *C. roseus* transformed with the halogenase genes²⁰. One of the early biosynthetic enzymes, strictosidine synthase, cannot turn over 5-chlorotryptamine²¹ (**2b**). Therefore, when transforming *C. roseus* with *pyrH* and *rebF*, we also introduced a mutant of strictosidine synthase (STRvm) that can convert 5-chlorotryptamine to 10-chlorostrictosidine^{16,22} (**4b**). After a selection process, we cultivated the transformed root culture on standard Gamborg's B5 plant medium and monitored chlorinated alkaloids using liquid chromatography/mass spectrometry (LC–MS). We observed formation of chlorinated tryptophans **1a** and **1b** and chlorinated alkaloids in both the RebH–RebF and PyrH–RebF–STRvm hairy root lines (Fig. 2 and Supplementary Figs 6–15). These results indicate that RebH, PyrH and the partner reductase function productively in the plant cell environment, demonstrating that the flavin halogenases are highly transportable among kingdoms. Because chlorinated alkaloid production was observed in the transformed lines, we conclude that tryptophan decarboxylase can competently turn over halogenated tryptophan substrates *in vivo*.

Hairy roots transformed with RebH and RebF, which produce 7-chlorotryptophan, yielded a major chlorinated product at *m/z* 359 (Fig. 2a). An authentic standard of 12-chloro-19,20-dihydroakuammicine (**5a**) co-eluted with this compound. Natural products containing the akuammicine scaffold have a variety of pharmacological activities^{23–25}. Although the parent compound, 19,20-dihydroakuammicine (**5**) has been isolated in good yields from other plants²⁶, it is not a major alkaloid in *C. roseus* hairy root culture. However, when wild-type *C. roseus* cell lines were incubated with 7-chlorotryptamine, 12-chloro-19,20-dihydroakuammicine was also the major chlorinated product (Supplementary Fig. 16). Therefore, the predominance of 12-chloro-19,20-dihydroakuammicine in the RebH–RebF hairy root line is probably due to substrate specificity of downstream enzymes for 7-chlorotryptamine. A hairy root line transformed with the 5-chlorotryptophan enzyme system, PyrH, RebF and STRvm, produced a variety of chlorinated alkaloids (Fig. 2b–d). Two representative chlorinated alkaloids, 10-chloroajmalicine (**6b**) and 15-chlorotabersonine (**7b**), were identified by co-elution with authentic standards²¹.

Chlorinated alkaloid production seemed to be stable over the course of at least six subcultures. The alkaloid 12-chloro-19,20-dihydroakuammicine was produced at 26 ± 4 μg per gram of fresh root weight of a representative cell line averaged over six subcultures. For comparison, wild-type cell lines produced ~ 25 μg per gram of fresh tissue weight of chlorinated alkaloids when the medium was supplemented with 200 μM 7-chlorotryptamine (**2a**). Similarly, 10-chloroajmalicine and 15-chlorotabersonine (**7b**) were produced at 2.8 ± 0.9 and 4.0 ± 1.0 μg per gram of fresh root weight, respectively, for a representative cell line averaged over four subcultures (Supplementary Figs

12 and 14). Different concentrations of KCl (3 μM –20 mM) were added to the medium, but increasing amounts of exogenous chloride salt did not significantly affect the yields of chlorinated alkaloids (Supplementary Figs 17 and 18).

Previous reports demonstrated that RebH can use bromide to yield brominated tryptophan⁸ (**1c**). To assess the capacity of RebH for bromination *in vivo*, we supplemented a low-chloride cell culture medium with KBr. The *in vitro* halide specificity of RebH correlated with the products generated *in vivo*, as we observed the formation of a compound that co-eluted with an authentic standard of 12-bromo-19,20-dihydroakuammicine (**5c**) (21 ± 8 and 49 ± 20 μg per gram of fresh root weight with 10 mM and 20 mM KBr supplementation, respectively; Fig. 3). In contrast, supplementation of the medium with KI failed to yield either iodinated tryptophan or iodinated alkaloids. Again, this correlated with *in vitro* studies showing that RebH does not accept iodide as a substrate⁸ (Supplementary Figs 19–22).

We also measured the transcript levels of the heterologous enzymes by real-time PCR with reverse transcription. The production of halogenated compounds depended on the expression of both RebF and RebH or PyrH. Notably, when the strictosidine synthase mutant STRvm was not expressed in the PyrH–RebF hairy root lines, we observed accumulation of 5-chlorotryptophan (representative cell line, 9 ± 1 μg per gram of fresh root weight) and 5-chlorotryptamine (representative cell line, 20 ± 9 μg per gram of fresh root weight), but did not observe downstream alkaloids (Supplementary Figs 23 and 24).

Tryptophan does not seem to accumulate in either wild-type or transformed hairy roots. However, accumulation of 7-chlorotryptophan (50 ± 12 μg per gram of fresh root weight for a representative RebH–RebF cell line) and 5-chlorotryptophan (8 ± 2 μg per gram of fresh root weight for a representative PyrH–RebF–STRvm cell line) was observed, suggesting that decarboxylation of chlorinated tryptophan is a bottleneck *in vivo*, a step that could potentially be subjected to future engineering efforts. This is consistent with the 30-fold-lower catalytic efficiency of the decarboxylase enzyme for halogenated

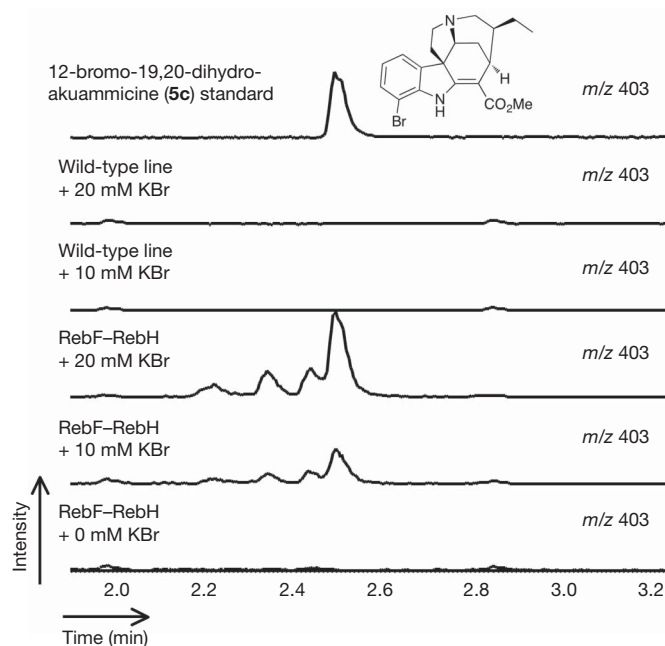


Figure 3 | Extracted LC–MS chromatograms showing the presence of 12-bromo-19,20-dihydroakuammicine (5c**; *m/z* 403) in RebF–RebH hairy roots.** Hairy roots are grown in medium supplemented with KBr (0–20 mM final concentration) for two weeks before alkaloid extractions. 12-bromo-19,20-dihydroakuammicine is not observed in control cultures transformed with no plasmid after incubation in KBr-supplemented medium. An authentic standard of 12-bromo-19,20-dihydroakuammicine is used to validate the structural assignment (Supplementary Figs 29, 30 and 32).

tryptophan *in vitro*. The morphologies of the halogen producing lines were thicker and slower growing than those of wild-type lines (Supplementary Fig. 25). Because tryptophan serves as the precursor for other small-molecule metabolites, we speculate that chlorinated tryptophan may be diverted into other pathways such as auxins. Notably, 4-chloro indole acetic acid, which is found in several species of pea, has altered activity relative to the auxin indole acetic acid^{27,28}.

Medicinal plants produce a wide range of complex natural products but generate relatively few halogenated compounds; chlorinated or brominated compounds are not found among the approximately 3,000 known monoterpene indole alkaloids produced by plants in the Apocynaceae, Rubiaceae and Loganiaceae families. The halogenation of natural products often has profound effects on the bioactivity of the compound, and can serve as a useful handle for further chemical derivatization^{1,29}. Despite the metabolic and developmental complexity of plant tissue, transformation of these prokaryotic genes led to the regioselective incorporation of halides into the alkaloid products of the existing plant pathway. Notably, the yield of chlorinated alkaloids in the most productive lines (~26 µg per gram of fresh weight of plant tissue) is only 15-fold lower than the yield of total natural alkaloids (compounds 5, 6, 7 and 8) from wild-type tissue (~420 µg per gram of fresh weight of plant tissue) (Supplementary Fig. 26). The ease with which we engineered the successful production of chlorinated alkaloids in *C. roseus*, a plant with limited genetic characterization, indicates that medicinal plants can provide a viable platform for synthetic biology.

METHODS SUMMARY

Structural characterization is shown in Supplementary Figs 27–32 and Supplementary Tables 1 and 2.

Generation of transgenic *C. roseus* hairy root cultures. We transformed the expression construct pCAMRebHRebF into *A. rhizogenes* ATCC 15834 by electroporation (1-mm cuvette, 1.25 kV), and we co-transformed pCAMPyrHRebF and pCAMSTRvm into *A. rhizogenes* ATCC 15834 by electroporation. Transformation of *C. roseus* seedlings with the generated *Agrobacterium* strains was performed as previously reported²⁰.

Evaluation of alkaloid production in transgenic *C. roseus* hairy roots. Every transgenic hairy root line that survived hygromycin selection medium was evaluated for alkaloid production. Transformed hairy roots were grown in Gamborg's B5 solid medium (half-strength basal salts, full-strength vitamins, 30 g l⁻¹ sucrose, 6 g l⁻¹ agar, pH 5.7). The total chloride concentration in Gamborg's B5 formulation was ~1 mM. We ground three-week-old hairy roots with a mortar, pestle and 106-µm acid-washed glass beads in methanol (10 ml g⁻¹ fresh weight of hairy roots). The crude natural product mixtures were filtered through 0.2-mm cellulose acetate membrane (VWR) and subsequently subjected to LC–MS analysis. Hairy roots transformed with wild-type *A. rhizogenes* lacking the plasmid were also evaluated.

Brominated alkaloid production in transgenic *C. roseus* hairy roots. We grew a selected transformed hairy root line for two weeks in low-chloride solid medium (67 mg l⁻¹ (NH₄)₂SO₄, 353 mg l⁻¹ Ca(NO₃)₂·4H₂O, 61 mg l⁻¹ MgSO₄, 1,250 mg l⁻¹ KNO₃, half-strength Murashige and Skoog micronutrient salts and full-strength Murashige and Skoog vitamins, 3 µM total chloride concentration). Hairy roots were transferred to the same medium supplemented with either potassium bromide or potassium iodide (10–20 mM final concentration) and cultivated for an additional two weeks. They were then processed and alkaloid production was analysed as described above (Supplementary Figs 12–15). We performed experiments in duplicate.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 April; accepted 20 September 2010.

Published online 3 November 2010.

1. Neumann, C. S., Fujimori, D. G. & Walsh, C. T. Halogenation strategies in natural products biosynthesis. *Chem. Biol.* **15**, 99–109 (2008).
2. Gribble, G. W. The diversity of naturally produced organohalogenes. *Chemosphere* **52**, 289–297 (2003).
3. Vaillancourt, F. H., Yeh, E., Vosburg, D. A., Garneau-Tsodios, S. & Walsh, C. T. Nature's inventory of halogenation catalysts: oxidative strategies predominate. *Chem. Rev.* **106**, 3364–3378 (2006).

4. van Pée, K. H. & Patallo, E. P. Flavin-dependent halogenases involved in secondary metabolism in bacteria. *Appl. Microbiol. Biotechnol.* **70**, 631–641 (2006).
5. Blasiak, L. C. & Drennan, C. L. Structural perspective on enzymatic halogenation. *Acc. Chem. Res.* **42**, 147–155 (2009).
6. Zehner, S. *et al.* A regioselective tryptophan 5-halogenase is involved in pyrroindomycin biosynthesis in *Streptomyces rugosporus* LL-42D005. *Chem. Biol.* **12**, 445–452 (2005).
7. Zhu, X. *et al.* Structural insights into regioselectivity in the enzymatic chlorination of tryptophan. *J. Mol. Biol.* **391**, 74–85 (2009).
8. Yeh, E., Garneau, S. & Walsh, C. T. Robust *in vitro* activity of RebF and RebH, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. *Proc. Natl Acad. Sci. USA* **102**, 3960–3965 (2005).
9. Yeh, E. *et al.* Flavin redox chemistry precedes substrate chlorination during the reaction of the flavin-dependent halogenase RebH. *Biochemistry* **45**, 7904–7912 (2006).
10. Yeh, E., Blasiak, L. C., Koglin, A., Drennan, C. L. & Walsh, C. T. Chlorination by a long-lived intermediate in the mechanism of flavin-dependent halogenases. *Biochemistry* **46**, 1284–1292 (2007).
11. Bitto, E. *et al.* The structure of flavin-dependent tryptophan 7-halogenase RebH. *Proteins* **70**, 289–293 (2008).
12. Sánchez, C. *et al.* Combinatorial biosynthesis of antitumor indolocarbazole compounds. *Proc. Natl Acad. Sci. USA* **102**, 461–466 (2005).
13. O'Connor, S. E. & Maresh, J. Chemistry and biology of terpene indole alkaloid biosynthesis. *Nat. Prod. Rep.* **23**, 532–547 (2006).
14. De Luca, V., Marineau, C. & Brisson, N. Molecular cloning and analysis of cDNA encoding a plant tryptophan decarboxylase: comparison with animal dopa decarboxylases. *Proc. Natl Acad. Sci. USA* **86**, 2582–2586 (1989).
15. McCoy, E. & O'Connor, S. E. Directed biosynthesis of alkaloid analogs in the medicinal plant *Catharanthus roseus*. *J. Am. Chem. Soc.* **128**, 14276–14277 (2006).
16. Bernhardt, P., McCoy, E. & O'Connor, S. E. Rapid identification of enzyme variants for reengineered alkaloid biosynthesis in periwinkle. *Chem. Biol.* **14**, 888–897 (2007).
17. Hawkins, K. M. & Smolke, C. D. Production of benzylisoquinoline alkaloids in *Saccharomyces cerevisiae*. *Nature Chem. Biol.* **4**, 564–573 (2008).
18. Minami, H. *et al.* Microbial production of plant benzylisoquinoline alkaloids. *Proc. Natl Acad. Sci. USA* **105**, 7393–7398 (2008).
19. De Luca, V. & Cutler, A. J. Subcellular localization of enzymes involved in indole alkaloid biosynthesis in *Catharanthus roseus*. *Plant Physiol.* **85**, 1099–1102 (1987).
20. Hughes, E. H., Hong, S.-B., Shanks, J. V., San, K.-Y. & Gibson, S. I. Characterization of an inducible promoter system in *Catharanthus roseus* hairy roots. *Biotechnol. Prog.* **18**, 1183–1186 (2002).
21. Rungtaphan, W. & O'Connor, S. E. Metabolic reprogramming of periwinkle plant culture. *Nature Chem. Biol.* **5**, 151–153 (2009).
22. Loris, E. A. *et al.* Structure-based engineering of strictosidine synthase: auxiliary for alkaloid libraries. *Chem. Biol.* **14**, 979–985 (2007).
23. Menzies, J. R., Paterson, S. J., Duwieuja, M. & Corbett, A. D. Opioid activity of alkaloids extracted from *Picralima nitida* (fam. Apocynaceae). *Eur. J. Pharmacol.* **350**, 101–108 (1998).
24. Fréderich, M. *et al.* Antiplasmodial activity of alkaloids from various strychnos species. *J. Nat. Prod.* **65**, 1381–1386 (2002).
25. Zhu, W.-M. *et al.* Components of stem barks of *Winchia calophylla* A. DC. and their bronchodilator activities. *J. Integr. Plant Biol.* **47**, 892–896 (2005).
26. Amat, M., Linares, A. & Bosch, J. A new synthetic entry to pentacyclic Strychnos alkaloids. Total synthesis of (+,–)-tubifolidine, (+,–)-tubifoline, and (+,–)-19,20-dihydroakumammicine. *J. Org. Chem.* **55**, 6299–6312 (1990).
27. Gandar, J. C. & Nitsch, C. Isolement de l'ester méthylique d'un acide chloro-3-indolylacétique à partir de graines immatures de pois, *Pisum sativum* L. *C. R. Acad. Sci. (Paris) Ser. D* **265**, 1795–1798 (1967).
28. Marumo, S., Hattori, H., Abe, H. & Munakata, K. Isolation of 4-chloroindolyl-3-acetic acid from immature seeds of *Pisum sativum*. *Nature* **219**, 959–960 (1968).
29. Deb Roy, A., Grischow, S., Cairns, N. & Goss, R. J. Gene expression enabling synthetic diversification of natural products: chemogenetic generation of pacidamycin analogs. *J. Am. Chem. Soc.* **132**, 12243–12245 (2010).
30. Li, S. *et al.* Assessment of the therapeutic activity of a combination of almitrine and raubasine on functional rehabilitation following ischaemic stroke. *Curr. Med. Res. Opin.* **20**, 409–415 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge support from the NIH (GM074820) and the American Cancer Society (RSG-07-025-01-CDD). We thank H.-Y. Lee and M. Tjandra for assistance with NMR characterizations and L. Li for high-resolution mass spectroscopy analysis.

Author Contributions All authors contributed to experimental design and data analysis. X.Q. initiated the project and its design, and performed steady-state kinetics. W.R. developed and implemented the transformation strategy and performed steady-state kinetics and metabolite analysis. All authors contributed to the preparation of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.E.O.C. (soc@mit.edu).

METHODS

Heterologous expression and purification of *C. roseus* tryptophan decarboxylase. The tryptophan decarboxylase (TDC) gene (accession number M25151.1) was obtained by reverse-transcription PCR (RT-PCR) amplification of mRNA isolated from *C. roseus* hairy root culture (Invitrogen, Dynabeads mRNA direct kit) with PCR primers that introduce sites for NdeI and XhoI (underlined): 5'-AAAAAACAATATGGGAGCATTGATTCACAA-3' and 5'-AAAAAAGCTCGA GTCAAGCTTCTTTGAGCAAATC-3'. The PCR fragment was subcloned into the pGEM-T Easy Vector (Promega), and then excised and ligated into the NdeI/XhoI site of the pET28a-plasmid (Novagen). The resulting pET28a-TDC construct was subsequently transformed into BL21 (DE3) pLysS electrocompetent *Escherichia coli* (Promega). A single *E. coli* colony harbouring pET28a-TDC was inoculated in 5 ml lysogeny broth medium supplemented with kanamycin (0.05 mg l⁻¹) and incubated overnight at 37 °C with shaking at 225 r.p.m. An aliquot of the overnight culture (1 ml) was then used to inoculate 100 ml lysogeny broth medium supplemented with kanamycin (0.05 mg l⁻¹) and incubated at 37 °C with shaking at 225 r.p.m. until OD₆₀₀ 0.6 was reached. Cells were induced for overexpression by the addition of isopropyl-β-D-galactopyranoside (IPTG; final concentration, 1 mM) and the culture was allowed to continue growth for 16 h at 18 °C. Cells were harvested by centrifugation and lysed by sonication. The hexahistidine-tagged TDC was purified using Ni-NTA Spin Kit (Qiagen) using manufacturer's protocols (Supplementary Fig. 1). Eluted enzyme was subsequently buffer-exchanged into phosphate buffer (50 mM NaH₂PO₄, 100 mM NaCl, pH 8.0) and immediately assayed for activity. This enzyme was not stable after extended storage.

Determining the steady-state kinetic constants of TDC for tryptophan substrate analogues, 5- and 7-chlorotryptophan (1b and 1a). Steady-state kinetic constants of TDC for 5- and 7-chlorotryptophan (1b and 1a) (Amatek) were determined in phosphate buffer (0.1 M NaH₂PO₄, 3.5 mM β-mercaptoethanol, pH 8.5) containing 1 mM pyridoxal-5'-phosphate at 30 °C (0.3-ml reaction volume) with TDC concentrations appropriate for obtaining the initial rate of the reaction (0.6–0.9 μM). Aliquots (25 μl) were quenched in 1 ml methanol, containing yohimbine (500 nM) as an internal standard, at appropriate time points. The samples were centrifuged (13,000 r.p.m. (16,000g), 5 min) to remove particulates and then analysed by LC-MS. Samples were ionized by ESI with a Micromass LCT Premier TOF Mass Spectrometer. The liquid chromatography was performed on an Acquity Ultra Performance BEH C18, 1.7 μm, 2.1 × 100 mm column on a gradient of 10–90% acetonitrile/water (0.1% formic acid) over 5 min at a flow rate of 0.6 ml min⁻¹. The appearance of the corresponding tryptamine analogues (either 5- or 7-chlorotryptamine) was monitored by peak integration and normalized to the internal standard. 5-chlorotryptamine was obtained from a commercial source (Alfa Aesar). 7-chlorotryptamine was synthesized as previously reported³. Eight substrate concentrations (200–2,500 μM) were tested for 7-chlorotryptophan substrate, and six substrate concentrations (200–1,200 μM) were tested for 5-chlorotryptophan substrate. Each concentration was assayed three times and the average values are reported with standard deviations. The data were fitted using nonlinear regression to the Michaelis–Menten equation using ORIGINPRO 7 (OriginLab). For reference, the kinetic constants for the natural substrate tryptophan (1) were also measured at concentrations ranging from 15 to 350 μM (Supplementary Fig. 2).

Construction of halogenase plant expression vectors PyrH–RebF–STRvm and RebH–RebF. The construction of plant expression vectors is summarized below and in Supplementary Fig. 3.

(i) The CaMV35S:Gus:NosPolyA fragment was obtained by PCR amplification of pCAMBIA1305.1 (Cambia) with forward and reverse PCR primers CaMV35S-NosPolyA that introduce sites for XbaI and KpnI at the 5' end, and PstI and SpeI at the 3' end (underlined): 5'-ACTTCTAGAGGTACCGGATCCTCTAGAGTCG ACCTGCAG-3' and 5'-ATTCTGCAGACTAGTCCCGATCTAGTAACATAG ATGACACCG-3'.

(ii) The tryptophan 5-halogenase gene (*pyrH*; accession number AAU95674) was obtained by PCR amplification of genomic DNA isolated from *Streptomyces rugosporus* NRRL 21084 with forward and reverse PCR primers CrPyrH that introduce sites for XhoI and NcoI at the 5' end, and SpeI and BstEII at the 3' end (underlined): 5'-ACTCTCGAGCCATGGATATCCGATCTGTGGTATCG-3' and 5'-ACTACTAGTGGTAACCTCATTGGATGCTGGCGAGGTA-3'.

(iii) The flavin reductase gene (*rebF*; accession number BAC15756) was obtained by PCR amplification of genomic DNA isolated from *Lechevalieria aerocolonigenes* ATCC 39243 with forward and reverse PCR primers CrRebF that introduce sites for XhoI and NcoI at the 5' end, and SpeI and PmlI at the 3' end (underlined): 5'-ACTCTCGAGCCATGGATACGATCGAGTTCGACAGAC-3' and 5'-ACT ACTAGTACCGTGTATCCCTCCGGTGTCCACAC-3'.

(iv) The tryptophan 7-halogenase gene (*rebH*; accession number BAC15758) was obtained by PCR amplification of genomic DNA isolated from *Lechevalieria*

aerocolonigenes ATCC 39243 with forward and reverse PCR primers CrRebH that introduce sites for SpeI and NcoI at the 5' end, and SpeI and PmlI at the 3' end (underlined): 5'-AAGACTACTAGTCCATGGATTCGGCAAGATTGAC-3' and 5'-ACTACTAGTACCGTGTACGCGCCGTGCTGTTGCC-3'.

CaMV35S:Gus:NosPolyA, PyrH, RebH and RebF PCR fragments were individually ligated into pGEM-Teasy vector (Promega) to yield pGEMCaMV35S, pGEMPyrH, pGEMRebH and pGEMRebF, respectively.

(v) Codon-optimization for *C. roseus* was performed to ensure efficient expression of the prokaryotic halogenase and flavin reductase genes in plant cell culture. Using codon usage database software (<http://www.kazusa.or.jp/codon/>), the following codons were identified as occurring at low frequency (triplet, frequency per thousand): GCG, 4.8; CGG, 3.3; ACG, 4.0; TCG, 6.1; CCG, 6.0; and CCC, 6.7. Site-directed mutagenesis was performed using a Stratagene QuikChange Site-Directed Mutagenesis kit to replace rare codons with the most frequently occurring codons encoding the corresponding amino acids. Only codons that appeared within the first 300 nucleotides of the genes were subjected to mutagenesis.

The site-directed-mutagenesis primers (name, sequence) were as follows (the sites of mutation are underlined): PyrH-SDM-for1, 5'-GTGGGTGGTGGC ACTGCTGGCTGGATGACC-3'; PyrH-SDM-rev1, 5'-GGTCATCCAGCCAGC AGTGCCACCAACCCAC-3'; PyrH-SDM-for2, 5'-GACATGCGGCCGTATAC TACTGCTACCGCGATGAGCGCCGGC-3'; PyrH-SDM-rev2, 5'-GCCGGCG CTCATCGCGGTAGCAGTAGTGTACGCCGCATGTC-3'; RebH-SDM-for, 5'-CCCCAATCTGCAGACTGCTTCTTCGACTTCCTCGGA-3'; RebH-SDM-rev, 5'-TCCGAGGAAGTCAAGAAAGCAGTCTGCAGATTGGGG-3'; RebF-SDM-for1, 5'-ACCGCGCGCGATCAAGGGCTCTGATGAGCCTGTTTCCC-3'; RebF-SDM-rev1, 5'-GGGAAACAGGGCTCATCAGAGCCCTGTGATCGGCC CGGGT-3'; RebF-SDM-for2, 5'-CTCGTCTGCCTGAACAGGGCTAGCGGAA CGTTGCAC-3'; RebF-SDM-rev2, 5'-GTGCAACGTTCCGCTAGCCCTGTTC AGGCAGACGAG-3'.

(vi) pGEMCaMV35S was digested and ligated into the KpnI/EcoRI sites of pSP72 vector (Promega) to yield pSPCaMV35S.

(vii) pGEMRebH and pGEMRebF were digested and ligated into the NcoI/PmlI sites of pSPCaMV35S to yield pSPRebH and pSPRebF, respectively. Similarly, pGEMPyrH was digested and ligated into the NcoI/BstEII sites of pSPCaMV35S to yield pSPPyH. pSPRebF was then digested and ligated into the PstI site of pSP72 to yield pSPRebF-2.

(viii) pSPRebH and pSPPyH were digested and ligated into the XbaI/EcoRI sites of pSPRebF-2 to yield pSPRebHRebF and pSPPyHRebF, respectively. Finally, both pSPRebHRebF and pSPPyHRebF were digested and ligated into the SpeI site of pCAMBIA1300A to yield pCAMRebHRebF and pCAMPyrHRebF. pCAMBIA1300A was constructed by introducing an SpeI restriction into pCAMBIA1300 (Cambia). The site-directed-mutagenesis primers are (SpeI site underlined) 5'-CCCGCCTTCAGTTTAACTAGTCACTGTTTGACAGGAT-3' and 5'-atcgtgtaaacactgACTAGTttaaactgaagcgagg-3'.

pCAMSTRvm was constructed as previously described²¹.

Generation of transgenic *C. roseus* hairy root cultures. The plant expression construct pCAMRebHRebF was transformed into *A. rhizogenes* ATCC 15834 by means of electroporation (1-mm cuvette, 1.25 kV). pCAMPyrHRebF and pCAMSTRvm were co-transformed into *A. rhizogenes* ATCC 15834 via electroporation (1-mm cuvette, 1.25 kV). Transformation of *C. roseus* seedlings with the generated *Agrobacterium* strains was performed as previously reported²¹. Briefly, 180–250 *C. roseus* seedlings (Vince Little Bright Eyes, Nature Hills Nursery) were germinated aseptically on Gamborg's B5 medium (full-strength basal salts, full-strength vitamins, 30 g l⁻¹ sucrose, pH 5.7) and grown in a 16-h light, 8-h dark cycle at 26 °C for 3 weeks. Seedlings were then wounded with extra-fine forceps at the stem tip, and 3–5 μl *A. rhizogenes* from a freshly grown liquid culture were inoculated on the wound.

Hairy roots appeared at the wound site 2–3 weeks after infection for about 80% of the seedlings infected. After hairy roots reached 1–4 cm in length (usually about 6 weeks after infection), they were excised and transferred to Gamborg's B5 solid medium (half-strength basal salts, full-strength vitamins, 30 g l⁻¹ sucrose, 6 g l⁻¹ agar, pH 5.7) containing hygromycin (0.03 mg ml⁻¹) for selection and the antibiotic cefotaxime (0.25 mg ml⁻¹) to remove remaining bacteria. The total chloride concentration in Gamborg's B5 formulation was 1 mM. All cultures were grown in the dark at 26 °C. After the hygromycin selection process, hairy roots were maintained in solid medium lacking both hygromycin and cefotaxime.

To adapt the line to liquid culture, approximately 200 mg of hairy roots (typically five 3–4-cm-long stem tips) from each line that grew successfully on solid medium were transferred to 50 ml of half-strength Gamborg's B5 liquid medium. The cultures were grown at 26 °C in the dark at 125 r.p.m. Hairy root growth in liquid medium seemed to be slower than that in solid medium. Hairy root transformants were screened for survival in solid medium supplemented with hygromycin. The number of transformants decreased significantly after solid medium

selection for each of the constructs transformed. Every line that grew in the selection medium was analysed for alkaloid production.

Hairy root selection and adaptation processes. For the plasmid pCAMRebH/RebF, the number of transformed hairy roots was 200 and the number of hairy roots after solid medium selection was 31. For the plasmid pCAMPyrH/RebF/STRvm, the number of transformed hairy roots was 140 and the number of hairy roots after solid medium selection was 57.

Verification of transferred DNA integration by genomic DNA analysis. To verify the integration of transferred DNA (T-DNA) into the plant genome, the genomic DNA from transformed hairy roots was isolated (Qiagen DNeasy kit) and then subjected to PCR amplification using T-DNA-specific primers with TDC primers serving as a positive control (see below). Specifically, for the pCAMRebH/RebF hairy roots, primers for PCR amplification were designed to amplify the complete TDC gene (TDC_for and TDC_rev), a 660-base-pair (bp) region of the RebH gene (RebH_for and RebH_rev), a 680-bp region of the RebF gene (RebF_for and RebF_rev) and an 800-bp region of the selection marker HPT gene (HPT_for and HPT_rev) (see below).

PCR primers for verification of T-DNA integration of transformed hairy roots were as follows (name, sequence): TDC_for, 5'-AAAAACATATGGGCAGCATTGATCAACA-3'; TDC_rev, 5'-AAAAAACTCGAGTCAAGCTTCTTTGAGCAATC-3'; RebH_for, 5'-GTCCTCGATGCCGACCTCTTC-3'; RebH_rev, 5'-GTACATGTCGATCTTCTCCTGC-3'; RebF_for, 5'-TAGAGGACCTAACAGAAC-3'; RebF_rev, 5'-CGTGACACTGGTCAGGGA-3'; HPT_for, 5'-GCCTGA ACTCACCGCGACGTC-3'; HPT_rev, 5'-CCTCCAGAAGAAGATGTTGGC-3'.

PCR amplification of genomic DNA from all of the selected transformed lines (pCAMRebH/RebF cell line 4, lanes 1–4; pCAMRebH/RebF cell line 5, lanes 5–8; pCAMRebH/RebF cell line 6, lanes 9–12; pCAMRebH/RebF cell line 10, lanes 13–16; and pCAMRebH/RebF cell line 11, lanes 17–20) was successful for all four sets of primers (Supplementary Fig. 4). PCR amplification of hairy root transformed with *A. rhizogenes* lacking the pCambia vector (provided by Professor Jacqueline Shanks (Iowa State University) and Professor Carolyn Lee-Parsons (Northeastern University)) genomic DNA was successful only when TDC-specific primers were used (lanes 21–24). These results indicated that *rebH* and *rebF* were successfully incorporated into the *C. roseus* genome in all chosen lines.

For the pCAMPyrH/RebF/STRvm hairy roots, primers for PCR amplification were designed to amplify the complete TDC gene (TDC_for and TDC_rev), the complete PyrH gene (PyrH_for and PyrH_rev), a 680-bp region of the RebF gene (RebF_for and RebF_rev), a 440-bp region of the STRvm gene and an 800-bp region of the selection marker HPT gene (HPT_for and HPT_rev) (see below and Supplementary Fig. 5).

PCR primers for verification of T-DNA integration of transformed hairy roots were as follows (name, sequence): TDC_for, 5'-AAAAACATATGGGCAGCATTGATCAACA-3'; TDC_rev, 5'-AAAAAACTCGAGTCAAGCTTCTTTGAGCAATC-3'; PyrH_for, 5'-ATGATCCGATCTGTGGTG-3'; PyrH_rev, 5'-TCATTGGATGCTGGCGAG-3'; RebF_for, 5'-TAGAGGACCTAACAGAAC-3'; RebF_rev, 5'-CGTGACACTGGTCAGGGA-3'; STRvm_for, 5'-CCTTATTATTGAAAGAGCTACATATG-3'; STRvm_rev, 5'-GCTAGAAACATAAGAAATTTCCCTTG-3'; HPT_for, 5'-GCCTGAACCTACCGCGACGTC-3'; HPT_rev, 5'-CCTCCAGAAGAAGATGTTGGC-3'.

PCR amplification of genomic DNA from three of four of the selected transformed lines (pCAMPyrH/RebF/STRvm cell line 1, lanes 1–5; pCAMPyrH/RebF/STRvm cell line 3, lanes 6–10; pCAMPyrH/RebF/STRvm cell line 6, lanes 11–15) was successful for all five sets of primers (Supplementary Fig. 5). PCR amplification of genomic DNA from pCAMPyrH/RebF/STRvm cell line 7 (lanes 16–20) was successful when TDC, PyrH, RebF and HPT primers were used but not when STRvm primers were used. PCR amplification of hairy root transformed with *A. rhizogenes* lacking the pCambia vector genomic DNA was successful only when TDC specific primers were used (lanes 21–25).

Evaluation of alkaloid production in transgenic *C. roseus* hairy roots. Every transgenic hairy root line that survived hygromycin selection medium was evaluated for alkaloid production. Transformed hairy roots were grown in Gamborg's B5 solid medium (half-strength basal salts, full-strength vitamins, 30 g l⁻¹ sucrose, 6 g l⁻¹ agar, pH 5.7). The total chloride concentration in Gamborg's B5 formulation was ~1 mM. Three-week-old hairy roots were ground with a mortar, pestle and 106 µm acid-washed glass beads in methanol (10 ml g⁻¹ of fresh hairy root weight). The crude natural product mixtures were filtered through 0.2-µm cellulose acetate membrane (VWR) and subsequently subjected to LC–MS analysis. Additionally, hairy roots transformed with wild-type *A. rhizogenes* lacking the plasmid were also evaluated.

These crude alkaloid mixtures were diluted 30:830 with methanol for mass spectral analysis. Samples were ionized by ESI with a Micromass LCT Premier TOF Mass Spectrometer. The liquid chromatography was performed on an Acquity Ultra Performance BEH C18, 1.7 µm, 2.1 × 100 mm column on a gradient

of 10–90% acetonitrile/water (0.1% TFA) over 5 min at a flow rate of 0.6 ml min⁻¹. The capillary and sample cone voltages were 1,300 and 60 V, respectively. The desolvation and source temperature were 300 and 100 °C, respectively. The cone and desolvation gas flow rates were 60 and 800 l per hour, respectively. Analysis was performed with MASSLYNX 4.1. Accurate mass measurements were obtained in W-mode. The spectra were processed using the MASSLYNX 4.1 mass measure, in which the mass spectrum of peaks of interest was smoothed and centred with TOF mass correction, locking on the reference infusion of reserpine. Data for RebF–RebH lines are shown in extracted LC–MS chromatograms in Supplementary Figs 6–10.

Feeding of 7-chlorotryptamine (2a) in control *C. roseus* hairy root cultures transformed with no plasmid. Alkaloid accumulation levels in hairy roots transformed with RebH and RebF were compared with alkaloid accumulation levels in control hairy root fed with 7-chlorotryptamine. The control hairy root line was grown for 2 weeks in half-strength Gamborg's B5 solid medium. Hairy roots were then transferred to the same medium supplemented with 7-chlorotryptamine (2a; 0, 25, 50, 100, 200 and 750 µM final concentrations) and grown for a further 1 week. Hairy roots were then processed and alkaloid production analysed as described in the previous subsection (Supplementary Fig. 16). Feeding studies were performed in duplicate.

Brominated alkaloid production in transgenic *C. roseus* hairy roots. A selected transformed hairy root line was grown for 2 weeks in low-chloride solid medium (67 mg l⁻¹ (NH₄)₂SO₄, 353 mg l⁻¹ Ca(NO₃)₂·4H₂O, 61 mg l⁻¹ MgSO₄, 1,250 mg l⁻¹ KNO₃, half-strength Murashige and Skoog micronutrient salts and full-strength Murashige and Skoog vitamins, 3 µM total chloride concentration). Hairy roots were transferred to the same medium supplemented with either potassium bromide or potassium iodide (10–20 mM final concentration) and cultivated for an additional 2 weeks. Hairy roots were then processed and alkaloid production analysed as described in the previous subsection but one. Hairy roots transformed with wild-type *A. rhizogenes* lacking the plasmid were also evaluated (Supplementary Figs 19–22). Experiments were performed in duplicate.

Purification and isolation of alkaloids from transformed TDC suppressed hairy roots supplemented with 7-chlorotryptamine and 7-bromotryptamine. To obtain chlorinated and brominated alkaloid standards, root tips (10–15) from TDC suppressed hairy roots³² were subcultured in six 50 ml Gamborg's B5 liquid medium (half-strength basal salts, full-strength vitamins, 30 g l⁻¹ sucrose, pH 5.7) and grown at 26 °C in the dark at 125 r.p.m. for 3 weeks before the medium was supplemented with either 7-chlorotryptamine (2a) or 7-bromotryptamine (2c) (750 µM final concentration). Both tryptamine analogue substrates were synthesized as previously reported³³. After 2 weeks of co-cultivation, hairy roots were extracted as described above in methanol (10 ml g⁻¹ fresh hairy root weight). Alkaloid extracts were filtered, concentrated under vacuum and redissolved in 25% acetonitrile/water (0.1% TFA) (1 ml g⁻¹ of fresh hairy root weight).

For cultures supplemented with 7-chlorotryptamine (2a), the redissolved mixture was purified on a 10 × 20 mm Vydec reverse-phase column using a gradient of 25–52% acetonitrile/water (0.1% TFA) over 24 min. Alkaloids were monitored at 228 nm and fractions containing the alkaloid analogues of interest, as determined by the characteristic isotopic distribution expected for chlorinated molecules (³⁵Cl/³⁷Cl) from LC–MS analysis, were combined and concentrated under vacuum (Supplementary Fig. 27).

For cultures supplemented with 7-bromotryptamine (2c), similar procedures were performed to isolate alkaloids from transgenic hairy roots, except that the liquid chromatography method was extended to 26 min. Alkaloids were monitored at 228 nm and fractions containing the alkaloid analogues of interest, as determined by LC–MS analysis, were combined and concentrated under vacuum (Supplementary Fig. 28).

Isolated alkaloids from both feedings were analysed by LC–MS (same parameters as above), analytical high-performance liquid chromatography and, where possible, high resolution LC–MS (Supplementary Table 1), ultraviolet–visible spectroscopy (Supplementary Fig. 29), tandem MS–MS (Supplementary Fig. 30) and ¹H NMR, ¹³C NMR and ¹H–¹³C HSQC using a Bruker AVANCE-600 NMR spectrometer equipped with a 5-mm 1H[13C,31P] cryoprobe (Supplementary Figs 31 and 32). Halogenated alkaloids generally displayed longer retention times than the natural alkaloids.

Quantification of chlorinated alkaloid production in transformed hairy roots. 12-chloro-19,20-dihydroakummicine (5a) and 12-bromo-19,20-dihydroakummicine (5c) standard curves were constructed by quantifying the peak areas of several concentrations (20–1,400 nM) of each alkaloid authentic standard using MASSLYNX 4.1. Similarly, 10-chloroajmalicine (6b), 15-chlorotabersonine (7b) and 10-chlorocatharanthine (8b) standard curves were constructed by quantifying the peak areas of several concentrations (20–1,400 nM) of each natural (that is, non-chlorinated) alkaloid authentic standard using MASSLYNX 4.1.

Dependence of chlorinated alkaloid production on concentrations of sodium chloride. A selected transformed hairy root line was grown for 2 weeks in low-chloride solid medium (67 mg l^{-1} $(\text{NH}_4)_2\text{SO}_4$, 353 mg l^{-1} $\text{Ca}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$, 61 mg l^{-1} MgSO_4 , $1,250 \text{ mg l}^{-1}$ KNO_3 , half-strength Murashige and Skoog micronutrient salts and full-strength Murashige and Skoog vitamins, $3 \mu\text{M}$ total chloride concentration). Hairy roots were transferred to the same medium supplemented with potassium chloride ($0\text{--}20 \text{ mM}$ final concentration) and grown for a further 2 weeks. Hairy roots were then processed and alkaloid production was analysed as previously described (Supplementary Figs 17 and 18).

Assessment of the stability of chlorinated alkaloid production in subsequent subcultures. Ten root tips from hairy roots transformed with pCAMRebH/RebF and pCAMPyrH/RebF/STRvm were subcultured every 3 weeks in Gamborg's B5 solid medium (half-strength basal salts, full-strength vitamins, 30 g l^{-1} sucrose, 6 g l^{-1} agar, pH 5.7), and grown at 26°C in the dark. Alkaloids were isolated from 21-day-old hairy roots and analysed as described above (Supplementary Figs 11–14).

Verification of expression of RebH, RebF and PyrH, STRvm enzymes by real-time RT-PCR. Real-time RT-PCR was used to assess the expression levels of RebH and RebF. Expression levels in hairy roots infected with *A. rhizogenes* lacking the pCAMRebH/RebF construct were compared with expression levels in hairy roots harbouring pCAMRebH/RebF. Messenger RNA from transformed hairy roots was isolated and purified from contaminant DNA using a Qiagen RNeasy Plant Mini Kit and Rnase-free DnaseI, respectively. The resulting mRNA was then reverse-transcribed to cDNA using a Qiagen QuantiTect Reverse transcription kit and then subjected to PCR with specific primers (see below), a Qiagen SYBR Green PCR kit and a Biorad DNA Engine Opticon 2 system. The threshold cycle (C_T) was determined as the cycle with a signal higher than that of the background plus 10 s.d. *Catharanthus roseus* 40S ribosomal protein S9 (Rps9), encoded by a house-keeping gene, was used to adjust the amount of the total mRNA in all samples. Real-time RT-PCR was performed in triplicate and the data are pictured as the relative expression levels of *rebH* and *rebF* mRNA in transgenic hairy roots as well as hairy roots lacking the pCAMBIA plasmid (Supplementary Fig. 23).

PCR primers for real-time RT-PCR of pCAMRebH/RebF transformed hairy roots were designed using the GenScript web tool (<http://www.genscript.com/ssl-bin/app/primer>), and were as follows (name, sequence): RebH_for, 5'-GACGG

GCATCTACTTCGTCT-3'; RebH_rev, 5'-TCGAACATCGTCTCGATCTC-3' (amplicon size, 117); RebF_for, 5'-CTGATGAGCCTGTTCCCA-3'; RebF_rev, 5'-CGTGACACTGGTCAGGGA-3' (amplicon size, 99); Rbps9_for, 5'-TTGAGC CGTATCAGAAATGC-3'; Rbps9_rev, 5'-CCCTCATCAAGCAGACCATA-3' (amplicon size, 122).

Real time RT-PCR was used to assess the expression levels of PyrH, RebF and STRvm. Expression levels in hairy roots infected with *A. rhizogenes* lacking the pCAMPyrH/RebF/STRvm construct were compared with expression levels in hairy roots harbouring pCAMPyrH/RebF/STRvm (Supplementary Fig. 24).

Chlorinated alkaloid production in a line lacking STRvm expression is shown in Supplementary Fig. 15. Photographs of the transformed roots are shown in Supplementary Fig. 25.

PCR primers for real-time RT-PCR of pCAMPyrH/RebF/STRvm transformed hairy roots were designed using GenScript web tool, and were as follows (name, sequence): PyrH_for, 5'-GCCTGCTCATCAACCAGAC-3'; PyrH_rev, 5'-CATC GCGGTAGCAGTAGTGT-3' (amplicon size, 137); RebF_for, 5'-CTGATGAG CCTGTTTCCCA-3'; RebF_rev, 5'-CGTGACACTGGTCAGGGA-3' (amplicon size, 99); STRvm_for, 5'-TATTATTGAAAGAGCTACATATG-3'; STRvm_rev, 5'-CTCTGCACTGCCTTTCTTG-3' (amplicon size, 134); Rbps9_for, 5'-TTGA GCCGTATCAGAAATGC-3'; Rbps9_rev, 5'-CCCTCATCAAGCAGACCATA-3' (amplicon size, 122).

Quantification of natural alkaloids in wild-type roots. The levels of natural alkaloids in wild type hairy roots was quantified as described in section on quantification of chlorinated alkaloid production in transformed hairy roots.

The levels of the four most abundant alkaloids found in these hairy roots, ajmalicine (6), tabersonine (7), catharanthine (8), as well as tryptophan (1) and tryptamine (2), were measured (Supplementary Fig. 26).

31. Hughes, E. H. *et al.* Characterization of an inducible promoter system in *Catharanthus roseus* hairy roots. *Biotechnol. Prog.* **18**, 1183–1186 (2002).
32. Runguphan, W. *et al.* Silencing of tryptamine biosynthesis for production of nonnatural alkaloids in plant culture *Proc. Natl Acad. Sci. USA* **106**, 13673–13678 (2009).
33. Schumacher, R. W. *et al.* Synthesis of didemnomolines A-D, N9-substituted β -carboline alkaloids from the marine ascidian *Didemnum* sp. *Tetrahedron* **55**, 935–942 (1999).

Design, function and structure of a monomeric ClC transporter

Janice L. Robertson¹, Ludmila Kolmakova-Partensky¹ & Christopher Miller¹

Channels and transporters of the ClC family cause the transmembrane movement of inorganic anions in service of a variety of biological tasks, from the unusual—the generation of the kilowatt pulses with which electric fish stun their prey—to the quotidian—the acidification of endosomes, vacuoles and lysosomes¹. The homodimeric architecture of ClC proteins, initially inferred from single-molecule studies of an elasmobranch Cl[−] channel² and later confirmed by crystal structures of bacterial Cl[−]/H⁺ antiporters^{3,4}, is apparently universal. Moreover, the basic machinery that enables ion movement through these proteins—the aqueous pores for anion diffusion in the channels and the ion-coupling chambers that coordinate Cl[−] and H⁺ antiport in the transporters—are contained wholly within each subunit of the homodimer. The near-normal function of a bacterial ClC transporter straitjacketed by covalent cross-links across the dimer interface and the behaviour of a concatemeric human homologue argue that the transport cycle resides within each subunit and does not require rigid-body rearrangements between subunits^{5,6}. However, this evidence is only inferential, and because examples are known in which quaternary rearrangements of extramembrane ClC domains that contribute to dimerization modulate transport activity⁷, we cannot declare as definitive a ‘parallel-pathways’ picture in which the homodimer consists of two single-subunit transporters operating independently. A strong prediction of such a view is that it should in principle be possible to obtain a monomeric ClC. Here we exploit the known structure of a ClC Cl[−]/H⁺ exchanger, ClC-ec1 from *Escherichia coli*, to design mutants that destabilize the dimer interface while preserving both the structure and the transport function of individual subunits. The results demonstrate that the ClC subunit alone is the basic functional unit for transport and that cross-subunit interaction is not required for Cl[−]/H⁺ exchange in ClC transporters.

To develop a strategy for generating a monomeric ClC protein, we examined the structure of ClC-ec1 (Fig. 1) for candidate residues mediating dimerization. This homologue is well suited to our purpose because its dimerization interface is almost completely membrane embedded, the large intracellular carboxy-terminal domain found in some ClC proteins being absent here. The interface is formed mainly by four helices running roughly perpendicular to the membrane to create a flat, nonpolar surface of $\sim 1,200 \text{ \AA}^2$ (Fig. 1). Most cross-subunit contacts are made by interdigitated leucine and isoleucine side chains; residues capable of forming hydrogen bonds or salt bridges are absent. The protein’s phospholipid-facing residues are also nonpolar (Fig. 1), a circumstance that invites questions of how such chemically similar surfaces so faithfully choose their respective protein and lipid partners in the dimer. Such questions have motivated extensive studies of transmembrane peptide dimerization^{8–10}, which identified shape complementarity as an important determinant of helix packing specificity within membranes and micelles. Shape complementarity of the ClC-ec1 dimer interface is high, scoring at levels seen for protein-antibody contacts and several membrane protein oligomers (Supplementary Table 1). Accordingly, our design strategy seeks to destabilize the dimer by placing steric mismatches on the ClC subunit interface. A second element of the strategy aims at favouring the interface’s exposure to the lipid bilayer. Lipid-facing surfaces of many membrane proteins are known to present amphiphilic tryptophan or tyrosine side chains to the chemically heterogeneous transition zone where the lipid acyl chains connect to the polar head groups, as seen for ClC-ec1 in Fig. 1; membrane-thermodynamic analysis of tryptophan analogues establishes that the aromatic, bifunctional character of this side chain favours its location at the phospholipid bilayer’s transition zone¹¹.

With these considerations in mind, we adopted a ‘warts-and-hooks’ strategy for engineering a monomeric ClC by introducing tryptophan

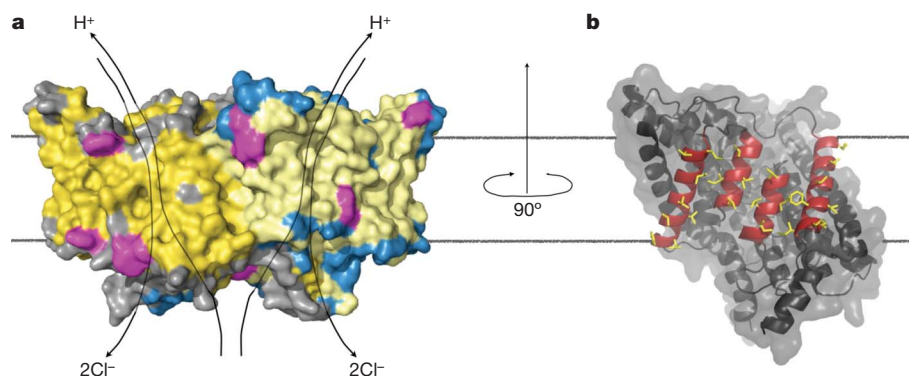


Figure 1 | Structure and dimeric interface of ClC-ec1. **a**, ClC-ec1 dimer (Protein Data Bank ID, 1OTS) is shown with subunits in grey and blue, with hydrophobic residues highlighted in yellow, and with tryptophan and tyrosine in magenta. The level of the membrane (extracellular side up) is indicated by black

lines. Previously proposed transport pathways are shown for Cl[−] and H⁺. **b**, Single subunit rotated 90° to view the dimerization interface head-on. The four interface helices (residues 192–204, 215–232, 405–416 and 422–440) are shown in red and the side chains involved in cross-subunit contacts are shown in yellow.

¹Department of Biochemistry, Howard Hughes Medical Institute, Brandeis University, Waltham, Massachusetts 02454, USA.

mutations on the subunit interface near the level of the lipid head groups. This type of substitution simultaneously offers two kinds of perturbation: steric disruption of the contact surface's shape complementarity and enhanced affinity of this surface for the lipid bilayer. We constructed eight single tryptophan substitutions for leucine or isoleucine near the extracellular and intracellular ends of the four dimerization helices (Fig. 2a). All but one of these mutants express near wild-type levels, and the oligomerization state of each was analysed in decylmaltoside micelles on a size exclusion column calibrated with a panel of membrane transport proteins¹² (Fig. 2b). The wild-type homodimer (100 kDa) elutes, as expected, at 12.8 ml, and the 50-kDa monomer is predicted to elute about 1 ml later. One mutant, Ile 422 Trp, shifts precisely to the presumed monomer position, with a minor dimer peak also apparent. Three other mutants, Ile 201 Trp, Leu 406 Trp and Leu 434 Trp, show broader, asymmetric peaks centred between dimer and monomer positions. The remaining three mutants all run as dimers (data not shown). In hopes of further stabilizing a monomer, we tested the double mutant Ile 201 Trp/Ile 422 Trp, which if dimeric would place four 'warts' within the subunit contact region, and if monomeric would offer two 'hooks' to the bilayer, one on each side of the membrane. This mutant, henceforth denoted WW, cleanly shifts to the monomer position with no observable dimer peak. The oligomeric nature of this double mutant in detergent micelles was further assessed by treatment with glutaraldehyde, a promiscuous crosslinker known quantitatively to produce covalent dimers of ClC-ec1¹³, as illustrated for the wild type by SDS–polyacrylamide gel electrophoresis (Fig. 2c). In contrast, glutaraldehyde treatment fails to shift WW to the covalent-dimer position, thereby identifying it as a monomer in detergent.

To identify the oligomeric state of WW in lipid bilayers, we repeated glutaraldehyde crosslinking experiments on this protein reconstituted into liposomes. Phosphatidylcholine–phosphatidylglycerol mixtures

were used here to avoid lipid-associated amino groups that would confound the glutaraldehyde reaction. We also aimed in these experiments to approximate Poisson-dilution conditions¹⁴, wherein a low protein/lipid ratio is used so that most liposomes are protein free and any liposome containing protein carries only a single transporting unit. Under such conditions, each liposome becomes a single-molecule reaction vessel in which intramolecular crosslinking is favoured. As shown in Fig. 3a, crosslinking in liposomes recapitulates the detergent results, thereby showing that WW is also monomeric in these bilayer membranes.

The experiments above establish the WW mutant as monomeric but do not address its conformational or functional character. We therefore performed two mechanistically diagnostic ion-transport measurements in the same liposome environment as was used for the crosslinking experiments. The unitary passive Cl[−] transport rate was determined in a 'Cl[−] dump' experiment¹⁴, in which liposomes with high Cl[−] concentration are suspended in low-Cl[−] solution in the presence of H⁺ and K⁺ ionophores, to prevent a pH gradient build-up and to maintain zero voltage. Under these conditions, the unitary Cl[−] efflux rate of wild-type protein, measured electrochemically by the appearance of Cl[−] in the external solution (Fig. 3b), is $\sim 300 \text{ s}^{-1}$; Cl[−] turnover by WW is roughly half of this value (160 s^{-1}), a respectable activity. Furthermore, anion specificity of transport is maintained in WW, as Cl[−] efflux is fully dependent on addition of K⁺ ionophore. ClC-ec1 is a coupled Cl[−]/H⁺ exchanger, in which a pre-established Cl[−] gradient can be used to pump H⁺ thermodynamically 'uphill'¹⁵. The WW monomer retains this defining feature of the transport

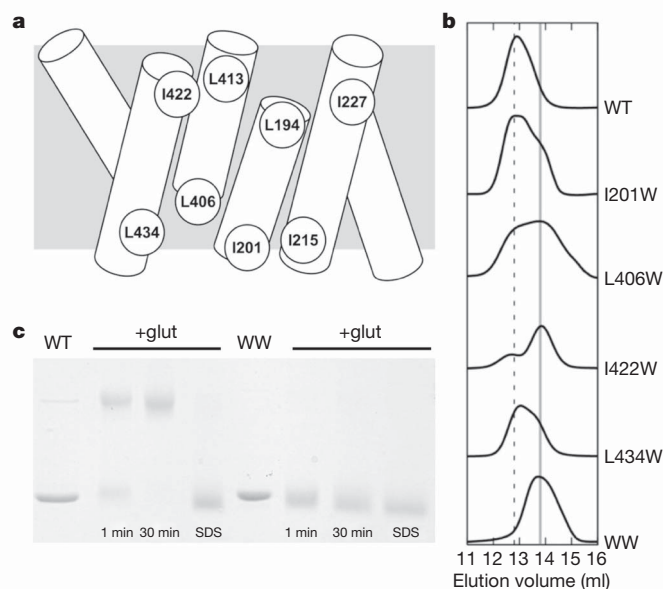


Figure 2 | Behaviour of tryptophan mutants in detergent. **a**, Schematic of the dimerization interface showing the positions of the tryptophans tested. Leu 194 Trp did not express protein. **b**, Chromatographic profiles of the various mutants on a Superdex 200 column. Vertical lines mark elution volumes for dimer (dashed) and monomer (solid). WT, wild type. **c**, 10% SDS–polyacrylamide gel electrophoresis of wild-type and WW samples, Coomassie stained. Bars indicate samples at 0.25 mg ml^{-1} treated with 0.125% glutaraldehyde, 150 mM NaCl and 50 mM Na phosphate, pH 7.0, for the indicated times in 5 mM decylmaltoside or, as a negative control, in 2% SDS. Crosslinking is nearly complete after 1 min, and no higher oligomers appear even after 30 min.

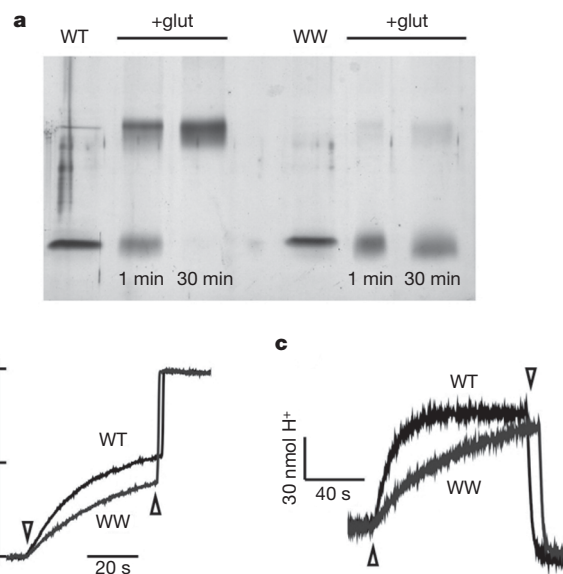


Figure 3 | Monomeric ClC mutant in phospholipid membranes.

a, Glutaraldehyde crosslinking of wild-type ClC-ec1 and the WW mutant in liposomes. Glutaraldehyde treatment was as in Fig. 2, except that protein was incorporated into phosphatidylcholine–phosphatidylglycerol liposomes, and gel was silver stained. **b**, Passive Cl[−] efflux from reconstituted liposomes for wild-type ClC-ec1 and the WW mutant. Traces show release of Cl[−] from liposomes loaded with 300 mM Cl[−] into the extraliposomal solution (containing 1 mM Cl[−]), initiated by 0.5 μM valinomycin (downward arrowhead), normalized to the level of complete release on disrupting liposomes with 50 mM octylglucoside (upward arrowhead). Unitary turnover calculated on a per-subunit basis from the initial rate of Cl[−] release¹⁴ was $290 \pm 30 \text{ s}^{-1}$ for wild type, $160 \pm 9 \text{ s}^{-1}$ for WW (mean \pm s.e.m., $N = 9$). [Cl[−]], Cl[−] concentration. **c**, Cl[−]-driven H⁺ pumping against a pH gradient. Liposomes loaded with 300 mM Cl[−], pH 5.0, were suspended in 1 mM Cl[−], pH 5.2, and transport was initiated by valinomycin (upward arrowhead) and terminated by carbonyl cyanide-*p*-trifluoromethoxyphenylhydrazone (FCCP, downward arrowhead), while the pH of the suspension was recorded. Upward deflection represents uptake of H⁺ into liposomes.

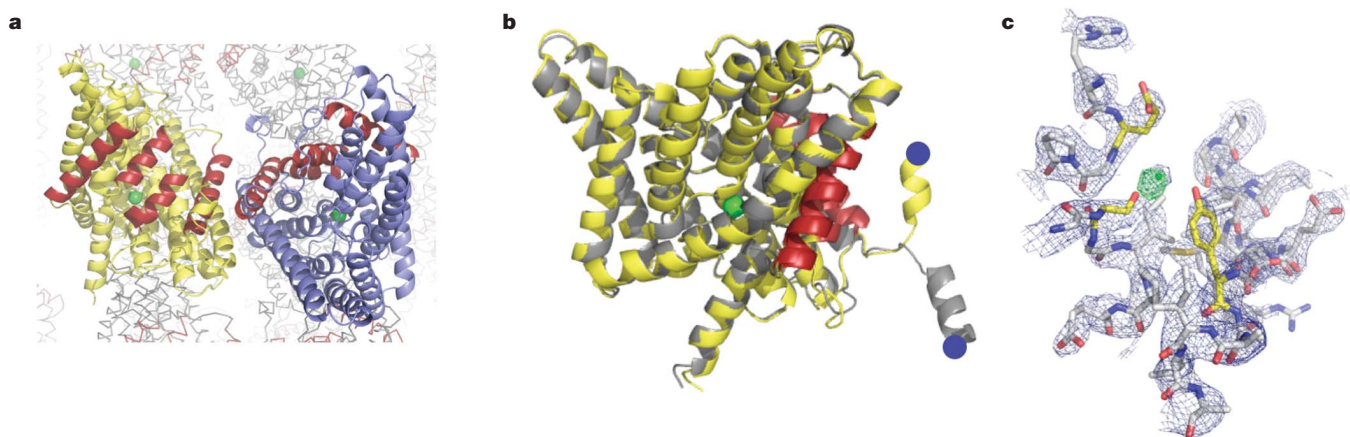


Figure 4 | Crystal structure of the WW monomer. **a**, View of two monomers in side-by-side contact, with interface helices highlighted in red, Cl^- ion highlighted in green and additional symmetry-related monomers shown in grey in the background. **b**, Backbone alignment ($\text{C}\alpha$ root mean squared deviation, 0.6 \AA) of the WW monomer (yellow, with interface helices in red) with a single subunit of wild-type CIC-ec1 (grey). Blue spheres indicate the

N termini of the visible structures. **c**, Central anion-binding site. The $2F_o - F_c$ map (blue, 1.5σ) is shown near the central Cl^- -binding site, with coordinating residues Ser 107, Glu 148 and Tyr 445 highlighted (yellow); the positive difference density calculated from a Cl^- -omit map (green) shows a strong peak (3.5σ) at the position of the central Cl^- ion in the wild type. Stereo versions of panels **a** and **c** can be found in Supplementary Fig. 2.

mechanism. As shown in Fig. 3c, Cl^- -loaded liposomes are suspended in low- Cl^- medium, and transport is initiated by depolarizing the liposomes with K^+ ionophore. As Cl^- flows out, H^+ enters against a pH gradient, as detected by alkalinization of the extraliposomal medium, which is swiftly reversed by addition of a proton ionophore. We established Cl^-/H^+ exchange stoichiometry from the ratio of initial flux rates (Supplementary Fig. 1): 2.0 ± 0.1 for the wild-type control, as expected from the two-to-one stoichiometry determined in *E. coli* lipids^{14–16}, and a similar value, 2.3 ± 0.3 , for WW. The preservation of H^+ -coupled Cl^- antiport in the monomeric construct directly establishes that the CIC subunit contains all essential components of the transport mechanism. The possibility remains that side-chain movements at the dimer interface in wild-type homodimer may occur during the transport cycle, as indicated convincingly by recent ^{19}F NMR experiments¹⁷, but our results demonstrate that such movements cannot represent functionally obligatory cross-subunit interactions.

We crystallized the WW mutant, collected X-ray diffraction data to a resolution of 3.1 \AA and solved the structure by molecular replacement using the wild-type subunit as search model (crystallographic statistics are shown in Supplementary Table 2). The asymmetric unit consists of a single monomer whose previously buried dimer interface is now completely exposed to detergent-containing solvent. This exposed interface is shown in Fig. 4a (also see Supplementary Fig. 2) for a symmetry-related pair of monomers, whose contacts in the unit cell arise from crystal geometry and are not seen in crystals of wild-type CIC-ec1. We consider it remarkable that the monomer's 18 membrane-embedded helices align precisely with those of the wild-type subunit in the homodimer (Fig. 4b), despite the absence of native cross-subunit interactions. Only the cytoplasmic amino-terminal helix (residues 22–30), which in the wild type engages in a domain swap with its twin subunit, veers off in a different direction to accommodate crystal packing. Moreover, most side chains projecting from the exposed subunit interface are well ordered and unperturbed from their buried positions in the wild-type dimer, except for a single tyrosine, which adopts a different rotamer to make room for one of the substituted tryptophans (Supplementary Fig. 3). Unambiguous density for the mechanistically crucial central Cl^- ion appears in the monomer at the same position as in the wild type, coordinated by the central serine and tyrosine residues (Fig. 4 and Supplementary Fig. 2); however, Cl^- density is lower in WW than in wild-type data sets of similar crystallographic quality¹⁸, perhaps because crystallization of the monomer requires the additional presence of NO_3^- , a transported anion known to compete with Cl^- (refs 13, 19).

A perplexing clash of form and function arises from this demonstration that the isolated CIC subunit is transport competent: why then are all known CIC proteins homodimers? With the steady expansion of the membrane protein structural database, it is becoming apparent that the parallel-pathways theme discussed here for CICs appears in many families of channels and transporters. For instance, aquaporin channels are homotetramers with a diffusion pore in each subunit²⁰, FNT-family formate channels are five-pore pentamers^{21,22} and UT-family urea channels²³, Amt-type ammonia channels^{24,25} and outer-membrane porins²⁶ are three-pore trimers. Among membrane transporters, a striking example is found in five phylogenetically unrelated families whose transporting subunits share a common structural fold but variously assemble as monomers, dimers or trimers^{27,28}. A survey of the current literature identifies no fewer than fourteen separate families ($\sim 40\%$ of structurally known membrane transport protein families) built on this parallel-pathway principle, with subunits held together through extended, nonpolar intramembrane contacts. We are loath to offer any suggestion for the ‘meaning’—evolutionary or physiological—of this emerging structural theme; in only one case, a trimeric Na^+ -coupled aspartate transporter of the EAAT superfamily²⁹, has parallel-pathway architecture been plausibly proposed as essential for substrate transport by the individual subunits making up the complex.

Although our warts-and-hooks design succeeded in severing the CIC dimer, we do not claim to understand the thermodynamic reasons for its success. The energetic components governing how a greasy protein surface chooses its greasy protein partner over a greasy lipid bilayer are still unparsed. Previous attempts to attack this fundamental problem of membrane protein chemistry have focused on model systems of single transmembrane helical peptides^{8–10,30}, and most have been quantifiable only in detergent micelles. The CIC interface introduced here may provide future opportunities to examine the molecular forces operating in transmembrane helix packing, folding and recognition in the context of a complex integral membrane protein.

METHODS SUMMARY

Expression in *E. coli*, purification and liposome reconstitution of CIC-ec1 (Swiss-Prot ID, P37019) were performed as described¹⁴, as were Cl^- and H^+ flux assays, except that we used lipid mixtures of egg phosphatidylcholine and 1-palmitoyl, 2-oleoyl phosphatidylglycerol in a 3/1 weight ratio. Ion flux rates in these lipids are 5–10-fold lower than observed in the *E. coli* phospholipids that we customarily use. Liposomes were formed at 20 mg ml^{-1} lipid, $1 \mu\text{g}$ protein per milligram lipid by dialysis, or by centrifugation of 0.1-ml samples through 3-ml Sephadex G-50 columns. Cl^-/H^+ exchange stoichiometry was determined as the ratio of initial

transport rates¹⁶, with Cl[−] efflux and H⁺ uptake recorded by means of Cl[−] and H⁺ electrodes using liposomes loaded with 300 mM KCl and 40 mM citrate-NaOH, pH 5.0, suspended in solutions of 1 mM KCl, 300 mM K isethionate and 2 mM citrate-NaOH, pH 5.2. For crystallography, 1 µl ClC protein (9–15 mg ml^{−1}) in 100 mM NaCl, ~40 mM decylmaltoside and 10 mM Tris-HCl, pH 7.5, was mixed with an equal volume of 100 mM LiNO₃, 41–45% (w/v) PEG400, 100 mM glycine-NaOH, pH 9.5, and ~10 mM 4-cyclohexyl-1-butyl-β-D-maltoside was added to the 2-µl drop. Crystals grown by vapour diffusion in sitting drop trays for 2–4 weeks at 20 °C were frozen in liquid nitrogen, and data were collected remotely at beamline 8.2.1 of the Advanced Light Source Eastern Annex, Waltham, Massachusetts. Data were processed in HKL2000. Molecular replacement was done in PHASER using residues 30–450 of a single subunit of ClC-ec1 as search model, and refinement was carried out in REFMAC5.

Received 22 June; accepted 5 October 2010.

Published online 3 November 2010.

- Jentsch, T. J. *et al.* Physiological functions of ClC Cl[−] channels gleaned from human genetic disease and mouse models. *Annu. Rev. Physiol.* **67**, 779–807 (2005).
- Middleton, R. E., Pheasant, D. J. & Miller, C. Reconstitution of detergent-solubilized Cl[−] channels and analysis by concentrative uptake of ³⁶Cl[−] and planar lipid bilayers. *Methods* **6**, 28–36 (1994).
- Dutzler, R. *et al.* X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* **415**, 287–294 (2002).
- Dutzler, R., Campbell, E. B. & MacKinnon, R. Gating the selectivity filter in ClC chloride channels. *Science* **300**, 108–112 (2003).
- Nguitragool, W. & Miller, C. ClC Cl[−]/H⁺ transporters constrained by covalent cross-linking. *Proc. Natl Acad. Sci. USA* **104**, 20659–20665 (2007).
- Zdebik, A. A. *et al.* Determinants of anion-proton coupling in mammalian endosomal ClC proteins. *J. Biol. Chem.* **283**, 4219–4227 (2008).
- Bykova, E. A. *et al.* Large movement in the C terminus of ClC-0 chloride channel during slow gating. *Nature Struct. Mol. Biol.* **13**, 1115–1119 (2006).
- Fleming, K. G., Ackerman, A. L. & Engelman, D. A. The effect of point mutations on the free energy of transmembrane α-helix dimerization. *J. Mol. Biol.* **272**, 266–275 (1997).
- MacKenzie, K. R. & Fleming, K. G. Association energetics of membrane spanning α-helices. *Curr. Opin. Struct. Biol.* **18**, 412–419 (2008).
- Chen, L. *et al.* Energetics of ErbB1 transmembrane domain dimerization in lipid bilayers. *Biophys. J.* **96**, 4622–4630 (2009).
- Yau, W. M. *et al.* The preference of tryptophan for membrane interfaces. *Biochemistry* **37**, 14713–14718 (1998).
- Fang, Y., Kolmakova-Partensky, L. & Miller, C. A bacterial arginine-agmatine exchange transporter involved in extreme acid resistance. *J. Biol. Chem.* **282**, 176–182 (2007).
- Maduke, M., Pheasant, D. J. & Miller, C. High-level expression, functional reconstitution, and quaternary structure of a prokaryotic ClC-type chloride channel. *J. Gen. Physiol.* **114**, 713–722 (1999).
- Walden, M. *et al.* Uncoupling and turnover in a Cl[−]/H⁺ exchange transporter. *J. Gen. Physiol.* **129**, 317–329 (2007).
- Accardi, A. & Miller, C. Secondary active transport mediated by a prokaryotic homologue of ClC Cl[−] channels. *Nature* **427**, 803–807 (2004).
- Miller, C. & Nguitragool, W. A provisional transport mechanism for a chloride channel-type Cl[−]/H⁺ exchanger. *Phil. Trans. R. Soc. B* **364**, 175–180 (2008).
- Elvington, S. M., Liu, C. W. & Maduke, M. C. Substrate-driven conformational changes in ClC-ec1 observed by fluorine NMR. *EMBO J.* **28**, 3090–3102 (2009).
- Accardi, A. *et al.* Synergism between halide binding and proton transport in a ClC-type exchanger. *J. Mol. Biol.* **362**, 691–699 (2006).
- Nguitragool, W. & Miller, C. Uncoupling of a ClC Cl[−]/H⁺ exchange transporter by polyatomic anions. *J. Mol. Biol.* **362**, 682–690 (2006).
- Fu, D. *et al.* Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* **290**, 481–486 (2000).
- Wang, Y. *et al.* Structure of the formate transporter FocA reveals a pentameric aquaporin-like channel. *Nature* **462**, 467–472 (2009).
- Waight, A. B., Love, J. & Wang, D. N. Structure and mechanism of a pentameric formate channel. *Nature Struct. Mol. Biol.* **17**, 31–37 (2010).
- Levin, E. J., Quick, M. & Zhou, M. Crystal structure of a bacterial homologue of the kidney urea transporter. *Nature* **462**, 757–761 (2009).
- Khademi, S. *et al.* Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 Å. *Science* **305**, 1587–1594 (2004).
- Zheng, L. *et al.* The mechanism of ammonia transport based on the crystal structure of AmtB of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **101**, 17090–17095 (2004).
- Cowan, S. W. *et al.* Crystal structures explain functional properties of two *E. coli* porins. *Nature* **358**, 727–733 (1992).
- Shaffer, P. L. *et al.* Structure and mechanism of a Na⁺-independent amino acid transporter. *Science* **325**, 1010–1014 (2009).
- Theobald, D. L. & Miller, C. Membrane transport proteins: surprises in structural sameness. *Nature Struct. Mol. Biol.* **17**, 2–3 (2010).
- Reyes, N., Ginter, C. & Boudker, O. Transport mechanism of a bacterial homologue of glutamate transporters. *Nature* **462**, 880–885 (2009).
- Metcalf, D. G. *et al.* Multiple approaches converge on the structure of the integrin αIIb/β3 transmembrane heterodimer. *J. Mol. Biol.* **392**, 1087–1101 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to the scientists at beamline 8.2.1 of the Advanced Light Source, Lawrence Berkeley Laboratory, for much help and advice, to H. Jayaram for help with refinement, and to R. Sah and D. Theobald for comments on the manuscript.

Author Contributions Experiments were designed by J.L.R. and C.M. and carried out by J.L.R. and L.K.-P., and the manuscript was written by all authors.

Author Information The atomic coordinates and structure factors of the WW (Ile201 Trp/Ile422 Trp) monomer have been deposited in the Protein Data Bank under accession code 3NMO. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.M. (cmiller@brandeis.edu).